UNIVERSITE DES ANTILLES ET DE LA GUYANE

U.F.R. Sciences de la Vie et de la Terre

THESE

pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DES ANTILLES ET DE LA GUYANE

Discipline : Sciences de la vie - Microbiologie

Ecole Doctorale Pluridisciplinaire de l'Université des Antilles Guyane

Présentée et soutenue publiquement par

Amal MOUMENE

Le 16 janvier 2015

Caractérisation de déterminants moléculaires du pouvoir pathogène d'*Ehrlichia ruminantium* : rôle du système de sécrétion de type IV et des protéines de la membrane externe

Co-Directeurs de thèse : Thierry LEFRANCOIS et Olivier GROS

Encadrant de thèse : Damien MEYER

Devant le jury composé de :

Matthieu **ARLAT**, Professeur de l'Université Paul Sabatier, Toulouse ;	Rapporteur
Matteo **BONAZZI**, Chargé de Recherche, HDR, CNRS, Montpellier ;	Rapporteur
Murielle **VAYSSIER-TAUSSAT**, Directrice de Recherche, Inra, Maison-Alfort ;	Examinatrice
Thierry **LEFRANCOIS**, Directeur Unité CMAEE, CIRAD, Montpellier ;	Examinateur
Olivier **GROS**, Professeur de l'Université Antilles-Guyane, Guadeloupe,	Examinateur
Damien **MEYER**, Chargé de Recherche, CIRAD, Guadeloupe ;	Examinateur

Recherches effectuées au sein de l'UMR CIRAD-INRA CMAEE,
«Contrôle des maladies animales exotiques et émergentes »,
Domaine de Duclos, 97170 Petit-Bourg, Guadeloupe

Financement de thèse : Union européenne, FEDER 1/1.4-30305, « Risques en santé animale et végétale » et CIRAD

AUTEUR : Amal Moumène

TITRE : Caractérisation de déterminants moléculaires du pouvoir pathogène d'*Ehrlichia ruminantium*: rôle du système de sécrétion de type IV et des protéines de la membrane externe.

DIRECTEURS DE THESE : Thierry Lefrançois / Olivier Gros

ENCADRANT : Damien Meyer

RESUME

Identifier les déterminants moléculaires des bactéries pathogènes et comprendre comment ils sont régulés pour permettre l'adaptation à l'environnement et à l'hôte est crucial pour imaginer des méthodes de contrôle innovantes et proposer des alternatives thérapeutiques. *Ehrlichia ruminantium* est une bactérie intracellulaire obligatoire de la famille des *Anaplasmataceae*, vectorisée par les tiques du genre *Amblyomma* et causant la cowdriose, une maladie fatale des ruminants. Dans le cadre de cette thèse, nous avons caractérisé certains déterminants du pouvoir pathogène d'*E. ruminantium* selon trois niveaux de résolution différents. Une approche globale sans *a priori*, nous a tout d'abord permis de déterminer le protéome de la membrane externe de la forme infectieuse d'*E. ruminantium*. Cette étude nous a permis d'avoir une meilleure vision de l'architecture de la membrane externe qui constitue la première interface d'échanges entre la bactérie et sa cellule hôte. Ensuite, nous avons montré la fonction d'ErxR comme régulateur central du système de sécrétion de type IV (SST4) et de la famille multigénique *map1* en permettant d'intégrer les signaux environnementaux que sont la carence en fer et l'acidité du milieu. En outre, ce travail a permis d'établir pour la première fois le lien entre les protéines Map1 de la membrane externe et le SST4 et suggère donc qu'elles puissent avoir un rôle direct dans la virulence. Enfin, une analyse *in silico* utilisant le logiciel S4TE a conduit à la caractérisation d'Erip1, le premier effecteur du SST4 d'*E. ruminantium*. Cet effecteur, phosphorylé sur les tyrosines et injecté dans le noyau de la cellule hôte, ne présente aucune homologie dans les bases de données et pourrait donc représenter une nouvelle famille de nucléomodulines. La recherche d'éventuels partenaires protéiques et l'identification des cibles intracellulaires de cet effecteur permettront de mieux comprendre comment *E. ruminantium* manipule la cellule hôte à son profit. Enfin, la caractérisation des voies de signalisation ciblées par Erip1 sera riche d'enseignements sur la réponse cellulaire à l'infection par *E. ruminantium*.

MOTS-CLES : *Ehrlichia ruminantium*, protéome, système de sécrétion de type IV, protéines de la membrane externe, Map1, ErxR, tr1, effecteurs, Erip1

DISCIPLINE : Microbiologie

SOMMAIRE

Avant-propos

Les maladies animales et zoonotiques émergentes provoquent chaque année de lourdes pertes parmi les élevages et présentent un risque croissant en termes de santé humaine pour les populations. Certaines bactéries pathogènes intracellulaires obligatoires de la famille des *Anaplasmataceae* sont les principaux agents des maladies émergentes transmises par les tiques (Rikihisa 2010). Comprendre les mécanismes moléculaires impliqués dans l'interaction hôte-pathogène, et plus spécifiquement l'adaptation à l'hôte et la mise en place du pouvoir pathogène chez des bactéries telles que *Ehrlichia* spp. ou *Anaplasma* spp. revêt un enjeu majeur pour la recherche de nouveaux traitements contre ces bactéries pathogènes. Au laboratoire, nous étudions la bactérie *Ehrlichia ruminantium*, responsable d'une maladie mortelle des ruminants, la cowdriose.

Alors que les bactéries du genre *Ehrlichia* subvertissent certains mécanismes majeurs de défense de l'hôte, les facteurs de virulence bactériens responsables de l'échec de ces défenses sont encore peu connus. Un des facteurs de virulence important pour l'infection intracellulaire chez les *Ehrlichia* est le système de sécrétion de type IV (SST4), qui délivre directement des protéines effectrices (ET4s) à l'intérieur des cellules hôtes eucaryotes. Ces ET4s ont été identifiés chez de nombreuses bactéries pathogènes *(Agrobacterium tumefaciens, Bartonella henselae, Brucella abortus, Anaplasma* spp., *Ehrlichia chaffeensis, Coxiella burnetii, Legionella pneumophila)* (Chen et al. 2010; Marchesini et al. 2011; Rikihisa and Lin 2010; Schulein et al. 2005; Vergunst et al. 2005; Zhu et al. 2011). Ils jouent un rôle crucial lors de l'infection et dans le développement de la maladie comme par exemple en inhibant la réponse immunitaire innée de l'hôte, en détournant les voies de signalisation cellulaire et le trafic vésiculaire ou en encore en utilisant les voies de recyclage liées à l'autophagie (Ivanov and Roy 2009; Newton, McDonough, and Roy 2013; Niu et al. 2012).

Plusieurs études ont révélé que le SST4 de nombreuses bactéries était sous le contrôle de mécanismes de régulation fins et variés (facteurs de transcription, signaux environnementaux) affectant l'activité des promoteurs des gènes codant le SST4 à différents degrés au cours du cycle de développement bactérien (Z. Cheng, Wang, and Rikihisa 2008; Wu et al. 2012). La compréhension globale et détaillée des voies de régulation du SST4 chez *E. ruminantium* en réponse à différents signaux environnementaux nous permettrait d'imaginer de nouvelles stratégies thérapeutiques en renforçant par exemple les voies perturbées par la bactérie par des approches pharmaceutiques.

Ma thèse a été menée au sein de l'unité « Contrôle des maladies animales exotiques et émergentes » au CIRAD en Guadeloupe et s'inscrit dans la thématique globale du laboratoire visant à étudier les interactions étroites entre la bactérie, la tique et l'hôte. Des approches sans *a priori* de type « omics » (transcriptomique, protéomique et métabolomique) sont actuellement développées au laboratoire mais des approches d'épidémio-surveillance sur la cowdriose et d'autres maladies animales sont aussi mises en œuvre au sein du réseau CaribVET coordonnée par l'équipe. L'intégration de l'ensemble des résultats permettra d'identifier des moyens de prévention et de lutte efficaces contre les différentes maladies prioritaires identifiées.

L'objectif principal de ma thèse était i) d'identifier certains déterminants moléculaires du pouvoir pathogène d'*Ehrlichia ruminantium* et ii) de comprendre leur régulation ou leur rôle dans l'adaptation d'*E. ruminantium* à son environnement, l'hôte animal étant un environnement particulier.

En guise d'introduction, une revue fait un état des connaissances sur les déterminants moléculaires du pouvoir pathogène chez les bactéries de la famille des *Anaplasmataceae,* i.e. *Ehrlichia* et *Anaplasma*. Ainsi, nous avons intégré et décrit l'ensemble des mécanismes d'infection de ces bactéries intracellulaires obligatoires en incluant la description des caractéristiques génomiques, l'entrée et la réplication dans la cellule eucaryote, la subversion des défenses de l'hôte, la manipulation de la machinerie cellulaire, la réplication intracellulaire, la lyse de la cellule hôte et la diffusion intercellulaire.

La partie résultats se compose de trois chapitres dont les travaux ont été menés en parallèle. Premièrement, nous présentons l'identification des protéines de la membrane externe d'*E. ruminantium* par une approche protéomique. Ensuite, nous détaillons l'étude de la régulation de l'expression des gènes codant pour le SST4 ou les protéines de la membrane externe de la famille Map1 en fonction des conditions environnementales et via une nouvelle protéine régulatrice majeure. Enfin, le dernier volet des résultats expose l'approche que nous avons menée pour identifier des effecteurs du SST4 d'*E. ruminantium* et qui a conduit à la caractérisation d'Erip1, un nouvel effecteur bactérien du SST4. La restitution des données est structurée sous la forme d'articles scientifiques soumis ou en cours de soumission pour publication. Enfin, les interrogations soulevées par les différents résultats obtenus, qui s'articulent autour des thèmes des facteurs de virulence, des protéines de la membrane externe et des effecteurs du SST4, seront discutés. Nous exposerons ensuite quelques perspectives suite à ce travail et proposerons les moyens à mettre en œuvre pour les réaliser.

INTRODUCTION GENERALE

Manuscrit préliminaire : *Ehrlichia* and *Anasplama*'s molecular tricks to manipulate their host

Amal Moumene[1,2,3] and Damien F. Meyer[1,2*]

[1]CIRAD, UMR CMAEE, Site de Duclos, Prise d'eau, F-97170 Petit-Bourg, Guadeloupe, France

[2]INRA, UMR1309 CMAEE, F-34398 Montpellier, France

[3]Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

* Correspondence to D.F.M, email: damien.meyer@cirad.fr

Ehrlichia and *Anaplasma*'s molecular tricks to manipulate their host cells

Amal Moumene[1,2,3] and Damien F. Meyer[1,2*]

[1]CIRAD, UMR CMAEE, Site de Duclos, Prise d'eau, F-97170 Petit-Bourg, Guadeloupe, France

[2]INRA, UMR1309 CMAEE, F-34398 Montpellier, France

[3]Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

* Correspondence to D.F.M, email: damien.meyer@cirad.fr

Abstract

Ehrlichia and *Anaplasma* are large genus of obligate intracellular Gram-negative bacteria transmited by ticks that cause several emerging infectious diseases in human, and are also pathogenic on rodents, ruminants, dogs and horses. *Ehrlichia* spp. and *Anaplasma* spp. invade and replicate either in endothelial cells or mammalian white blood cells and erythrocytes of the host or in midgut cells and salivary glands of the tick. In this review, we discuss the insights that functional studies are providing on how this group of bacteria exploits their host by subverting host innate immunity and hijacking cellular processes.

Introduction

Anaplasma spp. and *Ehrlichia* spp. are small Gram-negative obligately intracellular bacteria of the *Anaplasmataceae* family in the order *Rickttesiales* (Montagna et al., 2013). *Anaplasmataceae* amplify within host cell-derived vacuoles whereras *Rickettsiaceae* escape the host cell-derived vacuoles to grow free in the cytoplasm (Dumler and Walker, 2005). The capacity of these bacteria to invade and replicate inside host cells shows that they have developed strategies to perceive their environments and manipulate host cell functions to facilitate intracellular growth and spread infection.

Anaplasmataceae species were recognized to be mainly animal pathogens before they were reported to infect humans. The zoonotic pathogens *A. phagocytophilum, E. chaffeensis* and *E. ewingii* cause diseases called human granulocytic anaplasmosis (HGA), human monocytic ehrlichiosis (HGE) and *E. ewingii* ehrlichiosis, respectively (R. J. Thomas et al., 2009). These diseases induce symptoms that most commonly include fever, headache, myalgias, malaise and can be accompanied by thrombocytopenia, leukopenia, anemia and elevations in serum hepatic aminotransferase (Bakken et al., 1994). *E. canis* causes canine monocytic ehrlichiosis (CME) and few human cases have been reported (Perez et al., 2010). *E. ruminantium* is the causative agent of heartwater in ruminants and is suggested to be a potential emerging zoonotic pathogen, even if confirmation is needed (M. Allsopp et al., 2005).

In this review, we summarize recent advances in understanding the molecular mechanisms used by *Ehrlichia* spp. and *Anaplasma* spp. pathogenic bacteria to manipulate host cells, divert the cell machinery for their own profit and circumvent the immune system.

Life cycle and intracellular development

Ehrlichia spp. and *Anaplasma* spp. have the particularity to expand in two hosts, a tick vector and a mammalian host and have acquired highly sophisticated and diverse strategies to persist and infect their natural hosts (Rikihisa, 2010). Some species are associated with diverse host reservoirs including primary reservoirs like wild animals, deers (the most important reservoir in *E. chaffeensis*), rodents and domestic animals (e.g. dogs, ruminants, horses for *A. phagocytophilum*) serving

occasionally as secondary reservoirs for human infection (Ismail et al., 2010). These intracellular bacteria are transmitted between animals through the bites of infected ticks that are different according to species of bacteria (Rikihisa, 2010). These pathogens enter and replicate in different types of blood cells like granulocytes, monocytes, erythrocytes or platelets (B. A. Allsopp, 2010; De Tommasi et al., 2014). The diversity of vectors and reservoirs *of Anaplasmataceae* provides a wealth of information about life cycle and adaptation to various environments of these intracellular bacteria. *Ehrlichia* spp. and *Anaplasma* spp. display similar morphologies and life cycles (figures 1, 2 and 3). These organisms have a similar biphasic developmental cycle and transmission electron microscopy reveals two morphologically distinct forms (Zhang et al., 2007). First, the infectious extracellular forms (elementary bodies, EB or dense-core cells, DC) attach to the surface of host target cells before entering by endocytosis (figure 1). Inside the host cells, the bacteria develop within a membrane-bound vacuole where they differentiate into reticulate bodies (RB or reticulate cells, RC). They replicate to form a large colony, called morula, and after few days the bacteria redifferenciate into elementary bodies to be released outside the cell and to initiate a new infectious cycle (figure 1) (R. J. Thomas et al., 2009).

Ehrlichia **and** ***Anaplasma*** **spp. genome features**

Ehrlichia and *Anaplasma* species have a single, circular chromosome that vary in size from 0,9 Mb to 1,5 Mb for *E. ruminantium*, three times smaller than the genome of the intracellular bacteria *L. pneumophila* (Collins et al., 2005). The complete genome of most representative species of these genera have been sequenced, increasing the knowledge gained for the analyses and comparison of the genome sequences of different strains of *Anaplasmataceae* (Frutos, Viari, Ferraz, Bensaid, et al., 2006; Herndon et al., 2010; Thirumalapura et al., 2014). As a consequence of obligate intracellular lifestyles, genome reduction of *Anaplasmataceae* results in expendable genes due to the fact that some bacterial functions are accomplished by host (Collins et al., 2005; Mavromatis et al., 2006; Merhej and Raoult, 2011). A total of – 1000 proteins encoding genes are identified in *Ehrlichia* and *Anaplasma* species. The genomes of these bacteria shared conserved or unique genes but 15 to 36% of the genes have no homologies in the database (Dunning Hotopp et al., 2006).

Interestingly, the comparison of the genome sequences of three strains of *E. ruminantium* showed a high degree of genomic synteny and revealed the presence of specific features related to genomic plasticity, like high substitution rates, truncated genes and pseudogenes (Frutos, Viari, Ferraz, Bensaid, et al., 2006). Even if there are limited possibilities of gene acquirements via lateral transfer, ehrlichial genomes seem to evolve constantly, especially because of the presence of tandem repeats allowing reduction/expansion processes of these genomes (Frutos, Viari, Ferraz, Bensaid, et al., 2006). Proteins with tandem repeats (TRP) play important roles in pathogenicity such as interaction (Kumagai et al., 2010), actin nucleation (Jewett et al., 2006) and immune evasion (Gravekamp et al., 1996) in *Ehrlichia*. For instance TRP32, TRP47 and TRP120 secreted by T1SS (type I secretion system) of *E. chaffeensis* interact with genes and proteins host associated with signaling, vesicular trafficking and apoptosis (Wakeel et al., 2009; 2011; Luo and McBride, 2012) (figure 3). These TRPs proteins are differently expressed during the developmental cycle of the bacterium. The expression of TRP32 and TRP75 is constitutive in DC and RC whereas TRP47 and TR120 are expressed only at the late stages of infection (McBride et al., 2011). Interestingly, TRP orthologs p120/p140 of *E. chaffeensis* and *E. canis* elicit strong antibody responses, providing insight into the protective immune responses against these bacteria (Luo et al., 2009).

The analysis of the genome sequences for three *Anaplasma* species (*A. phagocytophilum*, *A. centrale*, and *A. marginale*) and three *Ehrlichia* species (*E. chaffeensis*, *E. canis*, and *E. ruminantium*) revealed their ability to synthetisize all nucleotides and most of vitamins and cofactors, whereas few amino acids are encoded by the genome. Some metabolic pathways such as systems for sugar uptake or glycolysis enzymes are limited in *A. marginale* (Brayton et al., 2005). In *E. ruminantium*, the glucose transport system is absent and the bacteria use proline and glutamate as primary source of carbone. Furthermore, the genomes of *E. ruminantium*, *A. phagocytophilum* and *E. canis* reveal an extensive aerobic respiratory chain (Collins et al., 2005). All these analyses suggest that certain metabolites and pathways seem to be fulfilled by the eukaryotic host cells, highlighting the strict dependency on host resources (Collins et al., 2005).

Sec–dependent and Sec-independent protein export pathways for secretion of proteins across the inner membrane are present in *Anaplasmataceae*. *E. chaffeensis* has a T1SS for the secretion of proteins with C-terminal secretion signal (Wakeel et

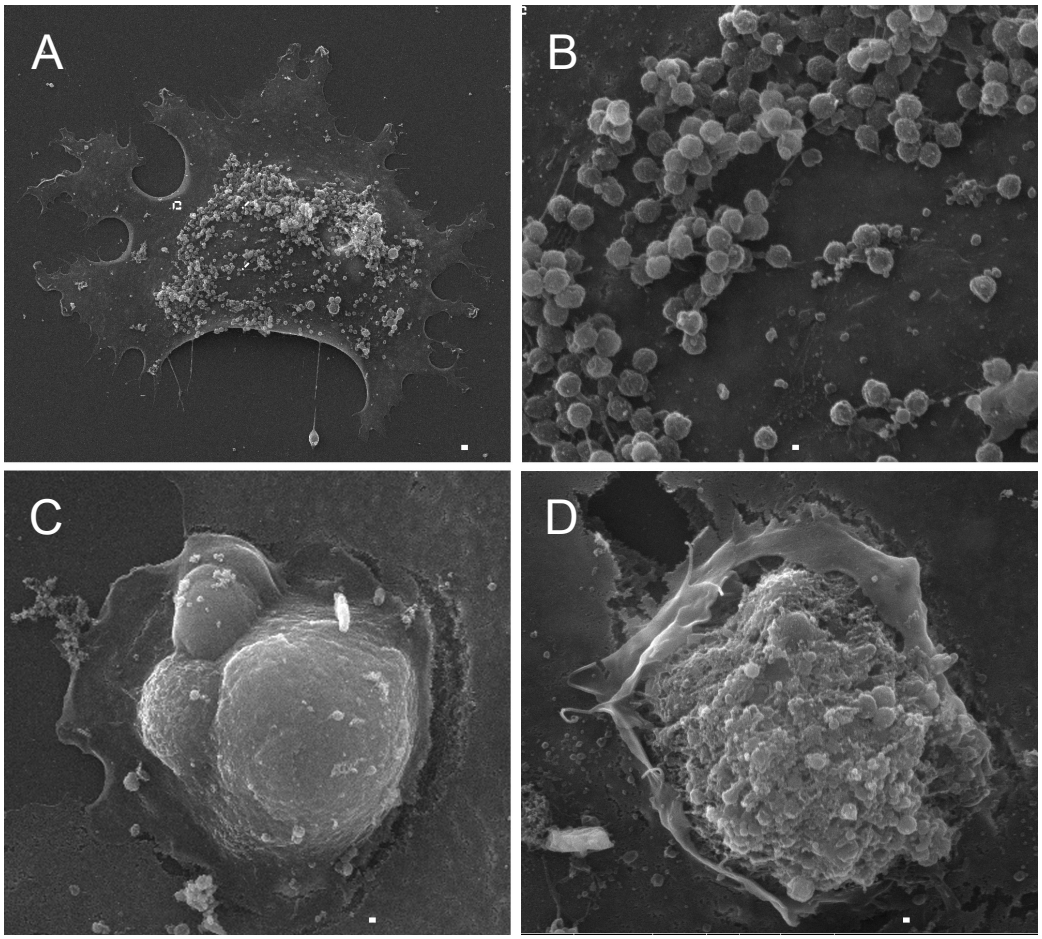


Figure 1. Scanning electron micrographs of numerous *Ehrlichia ruminantium* infecting a bovine endothelial cell. (A-B) First, *E. ruminantium* elementary bodies attach to the host cell surface. A zoom image of *E. ruminantium* bacteria adhering to host endothelial cell is shown in upper right corner (B). (C) Then, the bacteria get internalized and multiply inside the cell. (D) Infectious elementary bodies are released from infected host cell by complete cell lysis. Scale bar, 5 μm.

al., 2011). Furthermore, the type IV secretion (T4SS), which is a macromolecular complex that delivers protein substrates into the eukaryotic cells, is central to the pathogenesis of *Rickettsiales*. Among *Rickettsiales* order, the genetic diversity of T4SS is characterized by duplication of genes (e.g. *virB2*, *virB6*) and scattered genomic distribiution of the various operons encoding the T4SS and may be due to adaptations to various host cells (Gillespie et al., 2010; Al-Khedery et al., 2012). Paradoxically, the compact genomes of *Anaplasmataceae* harbor outer membrane proteines (OMP) encoding multigene families, namely the p44/msp2 superfamily. The orthologs are widely distributed in variable numbers within the genomes of *Ehrlichia* and *Anaplasma* spp. Although these outer membrane proteins are immunodominant proteins, whether they play role in bacterial adhesion or adaptation to host with a transport, function remains unknown. The evolutionary relationship between bacterial pathogens and their hosts is a continual process of adaptation, manifested by the gain and loss of genes via horizontal gene transfer (HGT). Le PT *et al.* revealed numerous HGT in *Rickettsiales* genomes, and thus permitting to draw the rhizome of *Rickettsiales* (Le et al., 2012).

Advances in functional genomics

The study of *Anaplasmataceae* pathogenesis has benefited from recent fundamental advances with genetic and functional genomic tools. Global transcriptome approaches are particularly useful to decipher the molecular mechanisms controlling bacterial pathogenesis.

Due to their obligate intracellular nature, the major limit for transcriptome analyses resides in the low quantity of prokaryotic mRNAs extracted from host cells and the contamination with eukaryotic mRNAs. Selective capture of transcribed sequences (SCOTS) was developed in *E. ruminantium* to capture bacterial mRNAs of *E. ruminantium* and permitted to identify genes related to the pathogenesis (Emboulé et al., 2009). For instance, *E. ruminantium* overexpresses genes involved in metabolism, nutrient exchange, and defense mechanisms, including those involved in resistance to oxidative stress at reticulate bodies stage suggesting that may undergoe oxidative stress and nutrient starvation conditions (Pruneau et al., 2012).

Moreover, a major technical hurdle in studying molecular determinants of *Anaplamataceae* pathogenesis has been removed with recent and rapid advances in

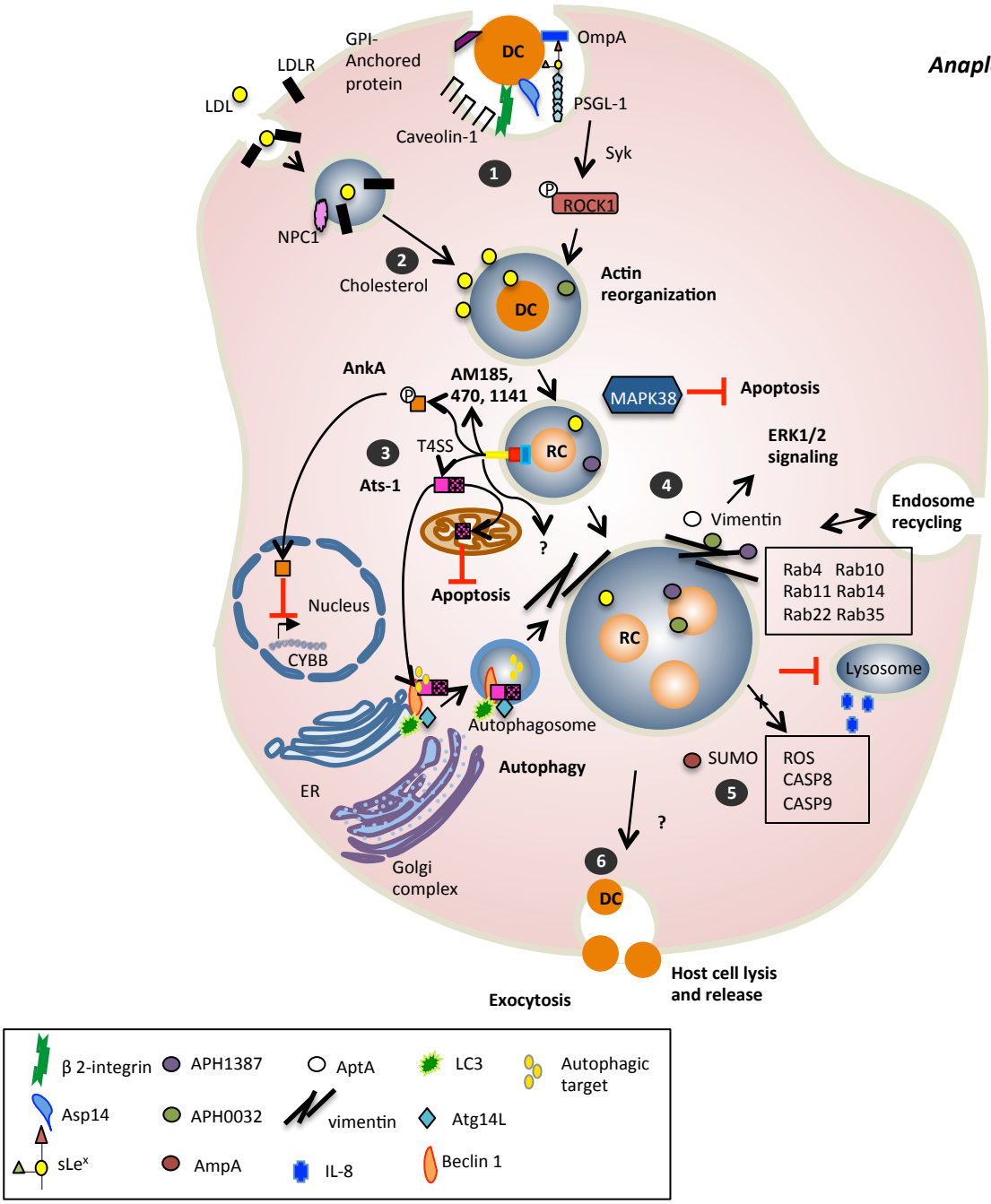


Figure 2. *Anaplasma* subverts host cell machineries to replicate inside host cell. (1) The DC enter and attach to mammalian host cell by using diverse outer membrane proteins such as OmpA, Aps14 that bind to PSGL-1 and the sLex tetrasaccharide. This leads to the phosphorylation of ROCK1 by Syk and allows bacterial uptake. (2) Bacteria acquire cholesterol by hijacking the LDLR pathway. (3) *Anaplasma* secretes two T4SS effectors in the host cytoplasm. AnkA is phosphorylated and translocated into the nucleus to regulate *CYBB* genes expression. After translocation into the cytoplasm, the full-length protein Ats-1 interacts with the host autophagosome initiation complex leading to autophagosome nucleation. A C-terminal portion of Ats-1 targets the mitochondria to inhibit apoptosis. (4) Vimentin surrounds the autophagosome and binds to AptA to modulate ERK1/2 signaling, contributing to intracellular survival. (5) The bacteria escape from the endosome–lysosome pathway and inhibit the production of ROS, CASP8 and CASP9. At the same time, AmpA interacts with SUMO from host cell to promote pathogen survival. (6) The DC are released by exocytosis or rupture of host cells. DC : Dense-cored cells; RC : Reticulate cells; ER : Endoplasmic reticulum; PSGL-1 : P-selectin glycoprotein ligand-1; T4SS : Type IV secretion system; SUMO : small ubiquitin-like modifier; ROCK1 : Rho-associated, coiled-coil-containing protein kinase 1; Syk : Spleen tyrosine kinase; LDLR : low density lipoprotein receptor; CYBB : Cytochrome b-245; ERK : Extracellular signal-regulated kinases; CASP : Caspase : ROS : Reactive oxygen species

genetic manipulation of several members of the *Anaplasmataceae* family. Thus, a random mutagenesis strategy using the Himar1 system has been applied to study *Anaplasma phagocytophilum*, *Anaplasma marginale*, *Ehrlichia chaffeensis* (Felsheim et al., 2006; 2010; C. Cheng et al., 2013). In *A. phagocytophilum*, transposon mutagenesis has disrupted a gene encoding a dihydrolipoamide dehydrogenase 1 (Ldpa1), APH_0065, an immunopathological molecule correlated with enhanced reactive oxygen species from NADPH oxidase and nuclear factor (NF)-κB signaling in macrophages (Chen et al., 2012). This result showed, for the first time in *A. phagocytophilum,* a connection between a mutation and its potential role during infection. Two recent mutagenesis experiments have been conducted in *A. marginale* by Pierle *et al*. to compare the transcriptional profiling between mutant and wild type strains. The authors identified gene mutations associated with some general metabolic pathways like nucleotide biosynthesis, translation elongation, correlated with a slow growth phenotype (Pierlé et al., 2013). These genes could represent good targets for the development of vaccines and new antibiotics. Another study demonstated that Himar1 transposon mutagenesis was achievable in *A. marginale* and led to the generation of a mutant inside an operon in the *omp10* gene, resulting in a significant decrease of transcripts and protein production of genes in this operon and altered virulence (Crosby et al., 2014). In *E. chaffeensis*, both targeted and random mutagenesis were developed and led to mutations in non-coding and coding regions (C. Cheng et al., 2013). The authors observed that mutations in certains genes inhibited infection of deers, natural reservoir of *E. chaffeensis*. The successful mutagenesis development of *Ehrlichia* and *Anaplasma* will permit to define genes important for pathogenesis, to verify the functions of candidate virulence factor genes and will open way for other related emerging zoonotic pathogens (C. Cheng et al., 2013). One of challenges will be to have the ability to cultivate the bacterium in axenic media, likewise in *Coxiella burnetii*, which has a similar developmental cycle (Omsland et al., 2009).

T4SS

Diverse bacterial pathogens use multiprotein complexes to deliver macromolecules into eukaryotic target cells to promote invasion and pathogenesis. The T4SS represents a major virulence determinant for *Anaplasmataceae* that harbor the prototypical *virBD4* T4SS of *A. tumefaciens*. In *A. phagocytophilum* and *E.*

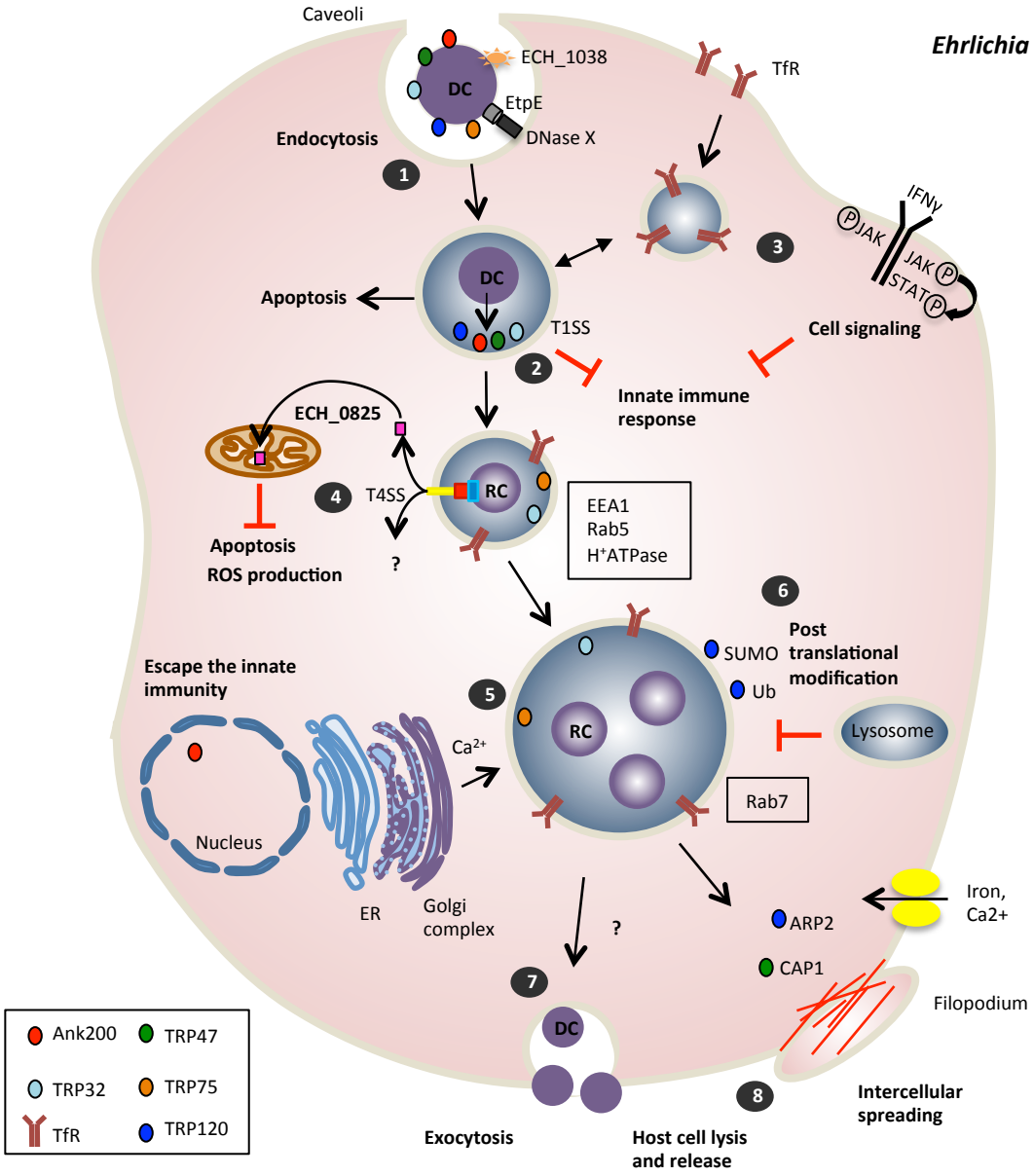


Figure 3. *Ehrlichia* develops highly adaptative strategies inside host cell. (1) The DC enter and attach to mammalian cells by using EtpE that binds to GPI-anchored protein DNase X. (2) The bacteria replicate in an ECV (*Ehrlichia*-containing vacuole) that resembles to early endosomes and secrete T1SS effector proteins (including TRP32, TRP47, TRP120 and Ank200) to escape host innate immune responses. (3) Then, the DC differentiate into RC. At this stage, the bacteria fuse with TfR endosome to acquire cholesterol and disrupt cell-signaling pathways like JAK/STAT to prevent innate immune response. (4) At the same time, *Ehrlichia* escapes the lysosomal pathway and secretes ECH_0825, a T4SS effector, to inhibit apoptosis and ROS production. (5) RC divide via binary fission to form microcolonies (morulae). (6) *Ehrlichia* exploits host SUMOylation pathways to mediate TRP120-host interactions to promote intracellular survival. (7) The DC are released by exocytosis or rupture of host cells. (8) *Ehrlichia* spread to neighboring cells through the host cell filopodium. DC : Dense-cored cells; RC : Reticulate cells; ER : Endoplasmic reticulum; T1SS : Type I secretion system; T4SS : Type IV secretion system; SUMO : small ubiquitin-like modifier; Ub : ubiquitination; JAK : Janus kinase; STAT : signal transducers and activators of transcription; TfR : Transferrin-receptor; CAP1 : Adenylate cyclase-associated protein 1; ARP2 : Actin-related proteins 2; system; IFNγ : Interferon gamma; ROS : Reactive oxygen species

chaffeensis, *virBD* genes are arranged in five clusters (Rikihisa, 2010). The conserved structure of splitted T4SS genomic islands and the duplicated *virB* genes suggests a common ancestral origin and an evolutionary pression for the persistence of duplicated genes (Gillespie et al., 2009; Rikihisa, 2010; Gillespie et al., 2010). Al-Khedery *et al*. showed a diversity of T4SS between the strains of *A. phagocytophilum* infecting different animals. For instance, *virB2* and *virB6* genes are present in different copies and have different sequences in each copy. This genetic diversity of the T4SS may be linked to differential virulence of *A. phagocytophilum* strains (Al-Khedery et al., 2012). The expression of *Anaplasma* spp. and *Ehrlichia* spp. T4SS components depends on the stage of the bacterial developmental cycle and the host. For instance, in *A. phagocytophilum, virB6* and *virB9* are up-regulated during infection of human neutrophil *in vitro*, while VirB9 is undetectable during the release of *A. phagocytophilum* (Niu et al., 2006). *E. chaffeensis* has four tandem paralogs *virB6* and *virB9* that are expressed in THP-1 cells and that etablish interaction with *Ehrlichia chaffeensis*-containing vacuoles (Bao et al., 2009). Moreover, the *virB2* paralogs identified in *A. phagocytophilum* are differentially transcribed between the human and tick cells (Nelson et al., 2008). The *virBD* genes are under a stringent regulation during the developmental cycle by transcription factors such as EcxR in *E. chaffeensis* (Z. Cheng et al., 2008). Some proteins of T4SS apparatus, such as VirB2, a putative VirB7, VirB11, and VirD4 in *A. marginale,* produce an immune response in infected or immunized animals and represent good targets for outer membrane vaccines (Sutten et al., 2010). In 2012, the Brown laboratory reported that VirB9-1, VirB9-2 and VirB10 are the highest immunogenic proteins, inducing IgG and stimulating CD4_T cells in *A. marginale* (Morse et al., 2012). These proteins are predicted to be surface exposed proteins and could be tested as vaccine candidates against *A. marginale* (Morse et al., 2012).

A functional T4SS and the related secretion of effectors (T4Es) is necessary for *Anaplasmataceae* to replicate inside their vacuole (Niu et al., 2012). In *A. phagocytophilum,* two T4Es are involved in pathogenesis (figure 2). The first effector secreted by T4SS is the protein AnkA that contains tandemly repeated ankyrin motifs (figure 2). This protein is tyrosine-phosphorylated in EPIYA motifs and adressed into the nucleus of host cell to decrease the *CYBB* gene expression (Garcia-Garcia et al., 2009; Zhu et al., 2009). This effector belongs to an emerging family of the nucleomodulins that manipulates host machinery (Bierne and Cossart, 2012). The

second effector of *A. phagocytophilum*, Ats-1 is targeted into the mitochondria of infected cells (figure 2). This protein can play two important roles: Ats-1 facilitates the recruitment of autophagosomes for the biogenesis of bacterial intracellular vacuole and Ats-1 is targeted into the mitochondria of infected cells to inhibit apoptosis (Niu et al., 2010). In *E. chaffeensis*, Ank200 is translocated in a T1SS-dependent manner using *E. coli* as a reporter system (figure 3) (Wakeel et al., 2011). This protein is translocated into the host cell nucleus where it could modify host cell signaling pathways to escape the innate immunity mechanisms (Zhu et al., 2009). *E. chaffeensis* T4SS is also used to insert bacterial effector proteins into host mitochondria. Using bacterial two-hybrid screening, the authors identified the first *ehrlichia* effector, ECH0825, involved in the inhibition of apoptosis to allow intracellular infection (figure 3) (H. Liu et al., 2012). Computational biology approaches coupled with experimental validation led to the identification of four T4SS effectors in *A. marginale* that were confirmed for their translocation in a T4SS-dependent manner (Lockwood et al., 2011). Moreover, Meyer *et al.* developed a bioinformatic algorithm for the prediction of T4 effectors from the genomes of *alpha-* and *gamma-proteobacteria*. This powerful tool will help for discovering new effectors involved in bacterial pathogenesis (Meyer et al., 2013).

Two-component systems

Many bacteria can sense changes in the environment by the mean of two component systems, a sensor kinase coupled with a response regulator, which allow a quick response via signal transduction. The genomes of *A. phagocytophilum* or *E. chaffeensis* encode three histidine kinases NtrX, PleC and CckA, that pair response regulators NtrY, PleD and CtrA respectively. PleD contains GGDEF domain protein associated with cyclic dimeric-GMP (c-di-GMP) activity (Römling, 2009; Lai et al., 2009). Expression of *pleD* and *pleC* occurs during exponential growth, at the morula stage and treatment with a hydrophobic c-di-GMP analog (CDGA) inhibits *A. phagocytophilum* infection (Kumagai et al., 2011). In *E. chaffeensis*, c-di-GMP controls TRP120 activity (involved in internalization of bacteria) and its signaling seems to be required for acute infection by this bacterium and contribute to virulence (Kumagai et al., 2010). Another response regulator, CtrA, is up-regulated in human monocytes during development of infectious DC form. LacZ reporter assays in *E. chaffeensis* showed that CtrA regulated positively *ompA* (peptidoglycan associated

lipoprotein), *bolA* (stress-induced morphogen) and *surE* (stationary-phase survival protein) and thus controls the stress resistance of DC (Z. Cheng et al., 2011).

Outer membrane proteins

The surface proteome of *A. marginale* infecting erythrocytes was studied by cross-linking reagent and revealed a covalent linkage of a group of surface exposed outer membrane proteins (OMPs) that has been tested to immunize cattle. This subset of outer membrane immunogens led to a significant level of protection. Differential expression of outer membrane proteins was revealed between mammalian and arthropod hosts, making them choice targets to block *Anaplasma* transmission (Noh et al., 2008). Recently, next-generation sequencing techniques showed a significantly genetic diversity in *A. marginale* subspecies *centrale* compared to that of US strains. This allowed the development of a multi-component recombinant vaccine (Dark et al., 2011). Cabezas-Cruz and co-workers analyzed the variation of the major surface protein 1a (MSP1a) of different strains of *A. marginale*. This immunogenic protein is an adhesin for bovine erythrocytes and tick cells. The study highlighted the potential biological implications of these genetic variations of sequences in key processes such as O-glycosylation, protein conformation, and pathogen-environmental interaction (Cabezas-Cruz et al., 2013). The most abundant OMP in *A. phagocytophilum* is Msp2 (P44), which seems to be involved in a mechanism of antigenic variation that facilitates persistent infection (Caspersen et al., 2002). Msp2 (P44) has been proposed to act as a porin (Wang et al., 2007) and may act as an adhesin to human granulocytes (Park et al., 2003). A proteomic approach showed that P44 and P28/OMP-1 are expressed in *A. phagocytophilum* and *E. chaffeensis* infected human leukocytes and thus suggested that these proteins must be important for the development of these bacteria and useful for investigation of novel antibacterial targets (Lin et al., 2011).

Host adaptation and subversion (figures 2 and 3)

The invasion process of *Anaplasma* and *Ehrlichia* species consists in four main steps: adhesion, internalisation, intracellular proliferation and intercellular spreading.

Adhesion

A. phagocytophilum interacts with host cell surface of neutrophils using adhesins carrying the N-terminal region of P-selectin glycoprotein ligand-1 (PSGL-1)

and sLex (figure 2) (Herron et al., 2000; Reneer et al., 2006). PSGL-1 and Syk proteins act cooperatively to phosphorylate ROCK1, allowing the bacterial uptake (figure 2) (V. Thomas and Fikrig, 2007). A comparison of the ability of DC and RC to bind to HL-60 cells or Chinese hamster ovary transfected to express PSGL-1 was performed in *A. phagocytophilum*. Monoclonal antibody against PSGL-1 reduced more significantly the binding of DC to PSGL-1 than for RC. This indicates that DC may be the appropriate form for the adhesion (Troese and Carlyon, 2009). Two proteins, OmpA and Asp14, interact with PSGL-1 to promote infection in mammalian host cells (Kahlon et al., 2013). The use of OmpA antiserum diminished infection of infect HL-60 cells by *A. phagocytophilum*, indicating the role of invasin to facilitate infection (Ojogun et al., 2012). Blocking Asp14 outer membrane protein interaction with the host cell inhibits infection of HL-60 cells by *A. phagocytophilum*. The combinatory effect of GST–Asp14 and GST-OmpA in mammalian host cells with led to a higher reduction of the infection. These combined proteins may serve as good targets against infection by multiple *Anaplasmataceae* pathogens (Kahlon et al., 2013). Moreover, lipid rafts, such as Caveolin-1, glycosylphosphatidylinositol anchored proteins (GAPs) and flotillin, facilitate the entry of *A. phagocytophilum* (figure 2) (Rikihisa, 2010). The bacterium also binds ß2-integrin, but the relevance of this interaction is only detectable under conditions mimicking the bloodstream (Schaff et al., 2010). In order to adhere and enter into mammalian cells, *E. chaffeensis* uses the invasin EtpE, an outer membrane protein, that binds to DNAse X, a GPI-anchored protein within caveoli at the monocyte cell surface (figure 3) (Mohan Kumar et al., 2013). Indeed, an antibody against EtpE inhibited interaction and infection.

Entry mechanisms and proliferation

A. phagocytophilum expresses some proteins inside its intracellular niche, termed *A. phagocytophilum*-occupied vacuole. This host cell-derived vacuolar membrane doesn't have early endosomal markers but contains late endosomal markers. The bacteria don't replicate in an acidified, lysosome-like vacuole (Huang, Troese, Howe, et al., 2010). APH_0032, APH_0233, and APH_1387 are associated with the formation of the vacuole. APH_1387 is expressed by RC form, whereas APH_0032 is induced by DC form (Huang, Troese, Ye, et al., 2010). Another protein, AptA decorates *A. phagocytophilum* inclusion and interacts with vimentin, the intermediate filament protein (Sukumaran et al., 2011). These results indicate that

Anaplasma is able to change its vacuolar membrane composition for efficient development. Some Rab proteins are sequestred by *A.phagocytophilum* to acquire amino acids and possibly cholesterol, and coat the vacuolar membrane in which the bacterium develops (figure 2). This aims at preventing maturation of the vacuole and lysosomal fusion. Rab proteins are involved in regulating endocytic recycling and transport of vesicles in the slow clathrin-independent pathway (Huang, Hubber, et al., 2010). Huang *et al.* demonstrated also that the vacuolar membrane *of A. phagocytophilum* is decorated by monoubiquitinated proteins (Huang et al., 2012). Similarly, following internalization, *E. chaffeensis* is contained in vacuoles that develop into early endosomes, which subsequently mature into late acidified endosomes but doesn't fuse with lysosomes. The early endosome show several hallmarks such as the early endosome antigen 1 (EEA1), the Rab5A protein, the transferrin receptor TfR and the vacuolar-type H+ ATPase (Barnewall et al., 1997; Mott et al., 1999). *E. chaffeensis* exhibits a late endosome characteristic, the protein Rab7 (Y. Cheng et al., 2014).

Subversion of host cell

Pathogen recognition and elimination of invading pathogens is essential for the control of bacterial infections. However, *Anaplasmataceae* developed strategies to bypass these processes, allowing them to survive and replicate intracellularly. Autophagy is the catabolic mechanism that detects and eradicates intracellular pathogens. Yet, some bacterial pathogens engage tactics to escape or inhibit autophagy and thus avoid lysosomal degradation (Baxt et al., 2013). Following entry into host cells, *A. phagocytophilum* subverts the endocytic pathway to avoid trafficking to lysosomes and thus replicates in a membrane-bound compartment (figure 2). This compartment does not show characteristics of endosomal or lysosomal markers but contains autophagosomal markers (Beclin 1, LC3, ATG6, ATG8) (Niu et al., 2008). A recent study confirmed that autophagy induced by the T4SS effector Ats-1 binding the host factor Beclin 1-Atg14L, allows the bacteria to acquire host nutrients for its replication (Niu et al., 2012). This suggests that *A. phagocytophilum* subverts autophagy to foster its own replication.

The effector AmpA (*A. phagocytophilum* post translationally modified protein A) and TRP120 are sumoylated on lysine residues during *A. phagocytophilum* and *E. chaffensis* infection, respectively (Beyer et al., 2014; Dunphy et al., 2014). In host

cells, these proteins colocalize with SUMO2/3 at the vacuole and in the cytosol. Treatment with anacardic acid known to disrupt SUMOylation reduced *A. phagocytophilum* load in infected cells and perturbed the interaction between TRP120 SUMOylation and known host proteins (Beyer et al., 2014; Dunphy et al., 2014). These studies show that *A. phagocytophilum* and *E. chaffeensis* also exploit host cell SUMOylation to promote their intracellular survival.

Ehrlichia spp. and *Anaplasma* spp. developed strategies to evade host innate defenses like apoptosis, to take advantage of the host cell. For instance, *E. ewingii* infection delays spontaneous apoptosis in host canine neutrophils *in vivo* by stabilizing mitochondrial membrane (Xiong et al., 2008). Similarly, *A. phagocytophilum* delays spontaneous human neutrophil apoptosis by inhibiting extrinsic pathway (inhibition of death receptor fas) and downstream intrinsic pathway (inhibition of caspase 8, Bax translocation, caspase 9 and XIAP degradation) (Ge and Rikihisa, 2006). *Ehrlichia* morulae interact with mitochondria to deliver proteins permitting the inhibition of mitochondrial activities (Y. Liu et al., 2011). It has been shown that Ank200, TRP32, TRP120 and the interaction between TRP47 and CAP1 promote also apoptosis during of infection of *E. chaffeensis* (Wakeel et al., 2009). During *Anaplasma* infection, the activation of p38 mitogen activated protein kinase (MAPK38) causes inhibition of apoptosis, activating expression of proinflammatory cytokines, (Choi et al., 2005). *A. phagocytophilum*-infected neutrophils induce the phosphatidylinositol 3-kinase/Akt pathway, allowing the release of IL-8 leading to apoptosis inhibition (Sarkar et al., 2012). As seen above, two effectors are translocated into mitochondria and involved in apoptosis inhibition. ECH0825 in *E. chaffeensis* is involved in curbing ROS and apoptosis (H. Liu et al., 2012). In *A. phagocytophilum*, Ats-1 induces the inhibition of apoptosis of host cells by preventing loss of mitochondrial membrane potential (Niu et al., 2010).

Moreover, the ERK/MAPK pathway is activated in *A. phagocytophilum*-infected human neutrophils (Lee et al., 2008). Interestingly, the drug manumycin A reduced the viability of *A. phagocytophilum* in host cells and ERK1/2 activation, making this drug therapeutic potential for HGA (Xiong and Rikihisa, 2011). AptA interacts with vimentin, a major constituent of the intermediate filament resulting in activation of ERK1/2. During infection, vimentin reorganized around the bacterial inclusion and may contribute to intracellular replication (Sukumaran et al., 2011). In

addition, *E. chaffeensis* can cause inflammatory responses through ERK pathways (Miura et al., 2011).

Anaplasma and *Ehrlichia* spp genomes do not encode for genes for the biosynthesis of lipid A, peptidoglycan or cholesterol (Collins et al., 2005; Brayton et al., 2005; Frutos, Viari, Ferraz, Morgat, et al., 2006). Instead, they subvert host cell cholesterol pathway and incorporate it into its outer membrane as a cell envelope constituent for survival and growth. *A. phagocytophilum* inclusion requires cholesterol from LDL uptake pathway, which is enhanced during infection of HL-60 cells (Xiong and Rikihisa, 2012). In addition, the expression of the low-density lipoprotein receptor is elevated. Thus, cholesterol accumulation seems to facilitate the growth of *A. phagocytophilum* (Xiong et al., 2009). The Niemann-Pick type C1 (NPC1) vesicule associated with LDL-derived cholesterol interacts with *A. phagocytophilum* inclusion for the acquisition of cholesterol. Blocking the NPC1 reduced significantly the infection by *A. phagocytophilum* (Xiong and Rikihisa, 2012).

A. phagocytophilum in host cells has the ability to stimulate the production of NADPH to scavenge exogenous superoxide ion O_2^- (Carlyon et al., 2004). *Ehrlichia* suppress reactive oxygen species (ROS) production, is highly sensitive to O_2^- and perturb Jak/STAT signaling, a process involved in the regulation of the immune system (Zhang et al., 2004). During infection, *E. chaffeensis* repressed immune responses like cytokine production (IL-12, IL-15, and IL-18) or activation of T_H1, NK cells, and cytotoxic T lymphocytes (Ismail et al., 2010; Mansueto et al., 2012).

Exit mechanisms and spreading

The mechanisms that mediate the release of *Ehrlichia* from cells are now better understood. *Ehrlichia* co-opts filopodia to traffic between cells during the initial stages of infection. Inhibition of filopodia formation by cytochalasin D prevents ehrlichial transport but the bacterium is then released by host cell membrane rupture adjacent to the morula during later stages of infection. This exit mechanism used by *Ehrlichia* allows the bacterium to evade the host innate immune system (S. Thomas et al., 2010). Similarly, inhibitory compounds affecting cytoskeleton re-arrangement, protein kinases, calcium channels or iron significantly reduced the number of *E. canis* in infected cells, indicating that these cellular processes are important for the proliferation of *E. canis* (Levenhagen et al., 2012). Alves *et al.* assessed the effect of several similar inhibitory drugs on spreading of *E. canis* in macrophages. They

showed that various host physiological processes like actin cytoskeleton, calcium and iron influx are required for full bacteremia and spreading in mammalian. Furthermore, acid phosphatase, used to label lysosomes, rarely marked the inclusions of *E. canis*, suggesting that *E. canis* escapes fusion with lysosome (Alves et al., 2014). Finally, TRP47 and TRP120 described above interact with host cytoskeletal proteins as well as with accessory proteins such as the ARP2/3 complex and CAP1 to facilitate exocytosis or filopodium formation (Dunphy et al., 2013).

Concluding statements

Considerable progresses have been made in understanding the pathogenesis of *Ehrlichia* and *Anaplasma* infection. Despite their obligate intracellular lifestyle, huge advances have been made in genetic manipulation of these bacteria. The recent development of targeted and random mutagenesis strategies, coupled with the tremendous evolution of sequencing technologies, omics approaches and *in vivo* imaging, offers new perspectives for the molecular dissection of the unique lifestyles and the virulence factors involved in the pathogenesis of these bacteria. Although many advances have been made in deciphering gene regulation of the intracellular growth and maturation of *Ehrlichia* and *Anaplasma*, many regulatory pathways still need to be discovered. Moreover, outstanding questions dealing with the ecology of these bacteria (e.g. the life inside the vector) or the determinants of host specificity remain largely unexplored but promise insightful knowledge on the biology of *Anaplasmataceae* and the adaptation to their host.

Acknowledgments

We would like to thank O. Gros of C3MAG microscopy plateform for his help in acquisition of SEM pictures. The authors acknowledge financial support from FEDER grant FED 1/1.4-30305, 2007–2013, “Risque en santé animale et végétale” (PhD grant to AM).

Figures

Figure.1 Scanning electron micrographs of numerous *Ehrlichia ruminantium* infecting a bovine endothelial cell

Figure. 2 *Anaplasma* subverts host cell machineries to replicate inside host cell

Figure. 3 *Ehrlichia* develops highly adaptative strategies inside host cell

References

Al-Khedery, B., Lundgren, A.M., Stuen, S., Granquist, E.G., Munderloh, U.G., Nelson, C.M., et al. (2012) Structure of the type IV secretion system in different strains of *Anaplasma phagocytophilum*. *bmc genomics* **13**: 678.

Allsopp, B.A. (2010) Natural history of *Ehrlichia ruminantium*. *Veterinary Parasitology* **167**: 123–135.

Allsopp, M., Louw, M., and Meyer, E.C. (2005) *Ehrlichia ruminantium*: an emerging human pathogen? *Annals of the New York Academy of Sciences* **1063**: 358–360.

Alves, R.N., Levenhagen, M.A., Levenhagen, M.M.M.D., Rieck, S.E., Labruna, M.B., and Beletti, M.E. (2014) The spreading process of Ehrlichia canis in macrophages is dependent on actin cytoskeleton, calcium and iron influx and lysosomal evasion. *Veterinary Microbiology* **168**: 442–446.

Bakken, J., Dumler, J., Chen, S., Eckman, M., Van Etta, L., and Walker, D. (1994) Human Granulocytic Ehrlichiosis in the Upper Midwest United States. A New Species Emerging? *The journal of the American medical association* **272**: 212–218.

Bao, W., Kumagai, Y., Niu, H., Yamaguchi, M., Miura, K., and Rikihisa, Y. (2009) Four VirB6 paralogs and VirB9 are expressed and interact in *Ehrlichia chaffeensis*-containing vacuoles. *Journal of Bacteriology* **191**: 278–286.

Barnewall, R.E., Rikihisa, Y., and Lee, E.H. (1997) *Ehrlichia chaffeensis* inclusions are early endosomes which selectively accumulate transferrin receptor. *Infect. Immun.* **65**: 1455–1461.

Baxt, L.A., Garza-Mayers, A.C., and Goldberg, M.B. (2013) Bacterial subversion of host innate immune pathways. *Science* **340**: 697–701.

Beyer, A.R., Truchan, H.K., May, L.J., Walker, N.J., Borjesson, D.L., and Carlyon, J.A. (2014) The *Anaplasma phagocytophilum* effector AmpA hijacks host cell SUMOylation. *Cell Microbiol* n/a–n/a.

Bierne, H. and Cossart, P. (2012) When bacteria target the nucleus: the emerging family of nucleomodulins. *Cell Microbiol* **14**: 622–633.

Brayton, K.A., Kappmeyer, L.S., Herndon, D.R., Dark, M.J., Tibbals, D.L., Palmer, G.H., et al. (2005) Complete genome sequencing of Anaplasma marginale reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 844–849.

Cabezas-Cruz, A., Passos, L.M.F., Lis, K., Kenneil, R., Valdés, J.J., Ferrolho, J., et al. (2013) Functional and immunological relevance of *Anaplasma marginale* major surface protein 1a sequence and structural analysis. *PLoS ONE* **8**: e65243.

Carlyon, J.A., Abdel-Latif, D., Pypaert, M., Lacy, P., and Fikrig, E. (2004) *Anaplasma phagocytophilum* utilizes multiple host evasion mechanisms to thwart NADPH oxidase-mediated killing during neutrophil infection. *Infect. Immun.* **72**: 4772–4783.

Caspersen, K., Park, J.-H., Patil, S., and Dumler, J.S. (2002) Genetic variability and stability of *Anaplasma phagocytophila msp2 (p44). Infect. Immun.* **70**: 1230–1234.

Chen, G., Severo, M.S., Sakhon, O.S., Choy, A., Herron, M.J., Felsheim, R.F., et al. (2012) *Anaplasma phagocytophilum* dihydrolipoamide dehydrogenase 1 affects host-derived immunopathology during microbial colonization. *Infect. Immun.* **80**: 3194–3205.

Cheng, C., Nair, A.D.S., Indukuri, V.V., Gong, S., Felsheim, R.F., Jaworski, D., et al. (2013) Targeted and random mutagenesis of *Ehrlichia chaffeensis* for the identification of genes required for *in vivo* infection. *PLoS Pathog* **9**: e1003171.

Cheng, Y., Liu, Y., Wu, B., Zhang, J.-Z., Gu, J., Liao, Y.-L., et al. (2014) Proteomic analysis of the *Ehrlichia chaffeensis* phagosome in cultured DH82 cells. *PLoS ONE* **9**: e88461.

Cheng, Z., Miura, K., Popov, V.L., Kumagai, Y., and Rikihisa, Y. (2011) Insights into the CtrA regulon in development of stress resistance in obligatory intracellular pathogen *Ehrlichia chaffeensis*. *Molecular Microbiology* **82**: 1217–1234.

Cheng, Z., Wang, X., and Rikihisa, Y. (2008) Regulation of type IV secretion apparatus genes during *Ehrlichia chaffeensis* intracellular development by a previously unidentified protein. *Journal of Bacteriology* **190**: 2096–2105.

Choi, K.S., Park, J.T., and Dumler, J.S. (2005) *Anaplasma phagocytophilum* delay of neutrophil apoptosis through the p38 mitogen-activated protein kinase signal pathway. *Infect. Immun.* **73**: 8209–8218.

Collins, N.E., Liebenberg, J., de Villiers, E.P., Brayton, K.A., Louw, E., Pretorius, A., et al. (2005) The genome of the heartwater agent Ehrlichia ruminantium contains multiple tandem repeats of actively variable copy number. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 838–843.

Crosby, F.L., Wamsley, H.L., Pate, M.G., Lundgren, A.M., Noh, S.M., Munderloh, U.G., and Barbet, A.F. (2014) Knockout of an outer membrane protein operon of *Anaplasma marginale* by transposon mutagenesis. *bmc genomics* **15**: 278–15.

Dark, M.J., Al-Khedery, B., and Barbet, A.F. (2011) Multistrain genome analysis identifies candidate vaccine antigens of *Anaplasma marginale*. *Vaccine* **29**: 4923–4932.

De Tommasi, A.S., Baneth, G., Breitschwerdt, E.B., Stanneck, D., Dantas-Torres, F., Otranto, D., and de Caprariis, D. (2014) *Anaplasma platys* in bone marrow megakaryocytes of young dogs. *J. Clin. Microbiol.* **52**: 2231–2234.

Dumler, J.S. and Walker, D.H. (2005) Rocky Mountain spotted fever--changing ecology and persisting virulence. *N. Engl. J. Med.* **353**: 551–553.

Dunning Hotopp, J.C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S.V., Eisen, J.A., et al. (2006) Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet* **2**: e21.

Dunphy, P.S., Luo, T., and McBride, J.W. (2014) *Ehrlichia chaffeensis* exploits host SUMOylation pathways to mediate effector-host interactions and promote intracellular survival. *Infect. Immun.* **82**: 4154–4168.

Dunphy, P.S., Luo, T., and McBride, J.W. (2013) *Ehrlichia* moonlighting effectors and interkingdom interactions with the mononuclear phagocyte. *Microbes Infect.* **15**: 1005–1016.

Emboulé, L., Daigle, F., Meyer, D.F., Mari, B., Pinarello, V., Sheikboudou, C., et al. (2009) Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium. BMC Mol. Biol.* **10**: 111.

Felsheim, R.F., Chávez, A.S.O., Palmer, G.H., Crosby, L., Barbet, A.F., Kurtti, T.J., and Munderloh, U.G. (2010) Transformation of *Anaplasma marginale*. *Veterinary Parasitology* **167**: 167–174.

Felsheim, R.F., Herron, M.J., Nelson, C.M., Burkhardt, N.Y., Barbet, A.F., Kurtti, T.J., and Munderloh, U.G. (2006) Transformation of *Anaplasma phagocytophilum*. *BMC Biotechnol.* **6**: 42.

Frutos, R., Viari, A., Ferraz, C., Bensaid, A., id, Morgat, A., Boyer, F., et al. (2006) Comparative genomics of three strains of *Ehrlichia ruminantium*: a review. *Annals of the New York Academy of Sciences* **1081**: 417–433.

Frutos, R., Viari, A., Ferraz, C., Morgat, A., Eychenié, S., Kandassamy, Y., et al. (2006) Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *Journal of Bacteriology* **188**: 2533–2542.

Garcia-Garcia, J.C., Rennoll-Bankert, K.E., Pelly, S., Milstone, A.M., and Dumler, J.S. (2009) Silencing of host cell *CYBB* gene expression by the nuclear effector AnkA of the intracellular pathogen *Anaplasma phagocytophilum*. *Infect. Immun.* **77**: 2385–2391.

Ge, Y. and Rikihisa, Y. (2006) *Anaplasma phagocytophilum* delays spontaneous human neutrophil apoptosis by modulation of multiple apoptotic pathways. *Cell Microbiol* **8**: 1406–1416.

Gillespie, J.J., Ammerman, N.C., Dreher-Lesnick, S.M., Rahman, M.S., Worley, M.J., Setubal, J.C., et al. (2009) An anomalous type IV secretion system in *Rickettsia* is evolutionarily conserved. *PLoS ONE* **4**: e4833.

Gillespie, J.J., Brayton, K.A., Williams, K.P., Diaz, M.A.Q., Brown, W.C., Azad, A.F., and Sobral, B.W. (2010) Phylogenomics reveals a diverse Rickettsiales type IV secretion system. *Infect. Immun.* **78**: 1809–1823.

Gravekamp, C., Horensky, D.S., Michel, J.L., and Madoff, L.C. (1996) Variation in repeat number within the alpha C protein of group B streptococci alters antigenicity and protective epitopes. *Infect. Immun.* **64**: 3576–3583.

Herndon, D.R., Palmer, G.H., Shkap, V., Knowles, D.P., and Brayton, K.A. (2010) Complete genome sequence of Anaplasma marginale subsp. centrale. *Journal of Bacteriology* **192**: 379–380.

Herron, M.J., Nelson, C.M., Larson, J., Snapp, K.R., Kansas, G.S., and Goodman, J.L. (2000) Intracellular parasitism by the human granulocytic ehrlichiosis bacterium through the P-selectin ligand, PSGL-1. *Science* **288**: 1653–1656.

Huang, B., Hubber, A., McDonough, J.A., Roy, C.R., Scidmore, M.A., and Carlyon, J.A. (2010) The Anaplasma phagocytophilum-occupied vacuole selectively recruits Rab-GTPases that are predominantly associated with recycling endosomes. *Cell Microbiol* **12**: 1292–1307.

Huang, B., Ojogun, N., Ragland, S.A., and Carlyon, J.A. (2012) Monoubiquitinated proteins decorate the *Anaplasma phagocytophilum*-occupied vacuolar membrane. *FEMS Immunol Med Microbiol* **64**: 32–41.

Huang, B., Troese, M.J., Howe, D., Ye, S., Sims, J.T., Heinzen, R.A., et al. (2010) *Anaplasma phagocytophilum* APH_0032 is expressed late during infection and localizes to the pathogen-occupied vacuolar membrane. *Microb. Pathog.* **49**: 273–284.

Huang, B., Troese, M.J., Ye, S., Sims, J.T., Galloway, N.L., Borjesson, D.L., and Carlyon, J.A. (2010) Anaplasma phagocytophilum APH_1387 is expressed throughout bacterial intracellular development and localizes to the pathogen-occupied vacuolar membrane. *Infect. Immun.* **78**: 1864–1873.

Ismail, N., Bloch, K.C., and McBride, J.W. (2010) Human Ehrlichiosis and Anaplasmosis. *Clinics in Laboratory Medicine* **30**: 261–292.
Jewett, T.J., Fischer, E.R., Mead, D.J., and Hackstadt, T. (2006) Chlamydial TARP is a bacterial nucleator of actin. *Proc. Natl. Acad. Sci. U.S.A.* **103**: 15599–15604.
Kahlon, A., Ojogun, N., Ragland, S.A., Seidman, D., Troese, M.J., Ottens, A.K., et al. (2013) *Anaplasma phagocytophilum* Asp14 is an invasin that interacts with mammalian host cells via its C terminus to facilitate infection. *Infect. Immun.* **81**: 65–79.
Kumagai, Y., Matsuo, J., Cheng, Z., Hayakawa, Y., and Rikihisa, Y. (2011) Cyclic dimeric GMP signaling regulates intracellular aggregation, sessility, and growth of *Ehrlichia chaffeensis*. *Infect. Immun.* **79**: 3905–3912.
Kumagai, Y., Matsuo, J., Hayakawa, Y., and Rikihisa, Y. (2010) Cyclic di-GMP signaling regulates invasion by *Ehrlichia chaffeensis* of human monocytes. *Journal of Bacteriology* **192**: 4122–4133.
Lai, T.-H., Kumagai, Y., Hyodo, M., Hayakawa, Y., and Rikihisa, Y. (2009) The *Anaplasma phagocytophilum* PleC histidine kinase and PleD diguanylate cyclase two-component system and role of cyclic Di-GMP in host cell infection. *Journal of Bacteriology* **191**: 693–700.
Le, P.T., Ramulu, H.G., Guijarro, L., Paganini, J., Gouret, P., Chabrol, O., et al. (2012) An automated approach for the identification of horizontal gene transfers from complete genomes reveals the rhizome of Rickettsiales. *BMC Evolutionary Biology* **12**: 243.
Lee, H.C., Kioi, M., Han, J., Puri, R.K., and Goodman, J.L. (2008) *Anaplasma phagocytophilum*-induced gene expression in both human neutrophils and HL-60 cells. *Genomics* **92**: 144–151.
Levenhagen, M.A., Alves, R.N., Rieck, S.E., Labruna, M.B., and Beletti, M.E. (2012) The role of cytoskeleton, components of inositol phospholipid signaling pathway and iron in *Ehrlichia canis in vitro* proliferation. *Vet. Res. Commun.* **36**: 195–199.
Lin, M., Kikuchi, T., Brewer, H.M., Norbeck, A.D., and Rikihisa, Y. (2011) Global proteomic analysis of two tick-borne emerging zoonotic agents: *anaplasma phagocytophilum* and *ehrlichia chaffeensis. Front. Microbio.* **2**: 24.
Liu, H., Bao, W., Lin, M., Niu, H., and Rikihisa, Y. (2012) *Ehrlichia* type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating host MnSOD. *Cell Microbiol* **14**: 1037–1050.
Liu, Y., Zhang, Z., Jiang, Y., Zhang, L., Popov, V.L., Zhang, J., et al. (2011) Obligate intracellular bacterium *Ehrlichia* inhibiting mitochondrial activity. *Microbes Infect.* **13**: 232–238.
Lockwood, S., Voth, D.E., Brayton, K.A., Beare, P.A., Brown, W.C., Heinzen, R.A., and Broschat, S.L. (2011) Identification of *Anaplasma marginale* type IV secretion system effector proteins. *PLoS ONE* **6**: e27724.
Luo, T. and McBride, J.W. (2012) *Ehrlichia chaffeensis* TRP32 interacts with host cell targets that influence intracellular survival. *Infect. Immun.* **80**: 2297–2306.
Luo, T., Zhang, X., and McBride, J.W. (2009) Major species-specific antibody epitopes of the *Ehrlichia chaffeensis* p120 and *E. canis* p140 orthologs in surface-exposed tandem repeat regions. *Clin. Vaccine Immunol.* **16**: 982–990.
Mansueto, P., Vitale, G., Cascio, A., Seidita, A., Pepe, I., Carroccio, A., et al. (2012) New insight into immunity and immunopathology of Rickettsial diseases. *Clin. Dev. Immunol.* **2012**: 967852.
Mavromatis, K., Doyle, C.K., Lykidis, A., Ivanova, N., Francino, M.P., Chain, P., et al. (2006) The genome of the obligately intracellular bacterium *Ehrlichia canis*

reveals themes of complex membrane structure and immune evasion strategies. *Journal of Bacteriology* **188**: 4015–4023.
McBride, J.W., Zhang, X., Wakeel, A., and Kuriakose, J.A. (2011) Tyrosine-phosphorylated *Ehrlichia chaffeensis* and *Ehrlichia canis* tandem repeat orthologs contain a major continuous cross-reactive antibody epitope in lysine-rich repeats. *Infect. Immun.* **79**: 3178–3187.
Merhej, V. and Raoult, D. (2011) Rickettsial evolution in the light of comparative genomics. *Biol Rev Camb Philos Soc* **86**: 379–405.
Meyer, D.F., Noroy, C., Moumène, A., Raffaele, S., Albina, E., and Vachiery, N. (2013) Searching algorithm for type IV secretion system effectors 1.0: a tool for predicting type IV effectors and exploring their genomic context. *Nucleic Acids Research* **41**: 9218–9229.
Miura, K., Matsuo, J., Rahman, M.A., Kumagai, Y., Li, X., and Rikihisa, Y. (2011) *Ehrlichia chaffeensis* induces monocyte inflammatory responses through MyD88, ERK, and NF-κB but not through TRIF, interleukin-1 receptor 1 (IL-1R1)/IL-18R1, or toll-like receptors. *Infect. Immun.* **79**: 4947–4956.
Mohan Kumar, D., Yamaguchi, M., Miura, K., Lin, M., Los, M., Coy, J.F., and Rikihisa, Y. (2013) *Ehrlichia chaffeensis* uses its surface protein EtpE to bind GPI-anchored protein DNase X and trigger entry into mammalian cells. *PLoS Pathog* **9**: e1003666.
Montagna, M., Sassera, D., Epis, S., Bazzocchi, C., Vannini, C., Lo, N., et al. (2013) "Candidatus Midichloriaceae" fam. nov. (Rickettsiales), an ecologically widespread clade of intracellular alphaproteobacteria. *Applied and Environmental Microbiology* **79**: 3241–3248.
Morse, K., Norimine, J., Palmer, G.H., Sutten, E.L., Baszler, T.V., and Brown, W.C. (2012) Association and evidence for linked recognition of type IV secretion system proteins VirB9-1, VirB9-2, and VirB10 in *Anaplasma marginale*. *Infect. Immun.* **80**: 215–227.
Mott, J., Barnewall, R.E., and Rikihisa, Y. (1999) Human granulocytic ehrlichiosis agent and *Ehrlichia chaffeensis* reside in different cytoplasmic compartments in HL-60 cells. *Infect. Immun.* **67**: 1368–1378.
Nelson, C.M., Herron, M.J., Felsheim, R.F., Schloeder, B.R., Grindle, S.M., Chavez, A.O., et al. (2008) Whole genome transcription profiling of *Anaplasma phagocytophilum* in human and tick host cells by tiling array analysis. *bmc genomics* **9**: 364.
Niu, H., Kozjak-Pavlovic, V., Rudel, T., and Rikihisa, Y. (2010) *Anaplasma phagocytophilum* Ats-1 is imported into host cell mitochondria and interferes with apoptosis induction. *PLoS Pathog* **6**: e1000774.
Niu, H., Rikihisa, Y., Yamaguchi, M., and Ohashi, N. (2006) Differential expression of VirB9 and VirB6 during the life cycle of *Anaplasma phagocytophilum* in human leucocytes is associated with differential binding and avoidance of lysosome pathway. *Cell Microbiol* **8**: 523–534.
Niu, H., Xiong, Q., Yamamoto, A., Hayashi-Nishino, M., and Rikihisa, Y. (2012) Autophagosomes induced by a bacterial Beclin 1 binding protein facilitate obligatory intracellular infection. *Proc. Natl. Acad. Sci. U.S.A.* **109**: 20800–20807.
Niu, H., Yamaguchi, M., and Rikihisa, Y. (2008) Subversion of cellular autophagy by *Anaplasma phagocytophilum*. *Cell Microbiol* **10**: 593–605.
Noh, S.M., Brayton, K.A., Brown, W.C., Norimine, J., Munske, G.R., Davitt, C.M., and Palmer, G.H. (2008) Composition of the surface proteome of *Anaplasma marginale* and its role in protective immunity induced by outer membrane

immunization. *Infect. Immun.* **76**: 2219–2226.
Ojogun, N., Kahlon, A., Ragland, S.A., Troese, M.J., Mastronunzio, J.E., Walker, N.J., et al. (2012) Anaplasma phagocytophilum outer membrane protein A interacts with sialylated glycoproteins to promote infection of mammalian host cells. *Infect. Immun.* **80**: 3748–3760.
Omsland, A., Cockrell, D.C., Howe, D., Fischer, E.R., Virtaneva, K., Sturdevant, D.E., et al. (2009) Host cell-free growth of the Q fever bacterium *Coxiella burnetii. Proc. Natl. Acad. Sci. U.S.A.* **106**: 4430–4434.
Park, J., Choi, K.S., and Dumler, J.S. (2003) Major surface protein 2 of *Anaplasma phagocytophilum* facilitates adherence to granulocytes. *Infect. Immun.* **71**: 4018–4025.
Perez, M., Bodor, M., Zhang, C., Xiong, Q., Xiong, Q., and Rikihisa, Y. (2010) Human infection with *Ehrlichia canis* accompanied by clinical signs in Venezuela. *Annals of the New York Academy of Sciences* **1078**: 110–117.
Pierlé, S.A., Hammac, G.K., Palmer, G.H., and Brayton, K.A. (2013) Transcriptional pathways associated with the slow growth phenotype of transformed *Anaplasma marginale*. *bmc genomics* **14**: 272.
Pruneau, L., Emboulé, L., Gely, P., Marcelino, I., Mari, B., Pinarello, V., et al. (2012) Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development. *FEMS Immunol Med Microbiol* **64**: 66–73.
Reneer, D.V., Kearns, S.A., Yago, T., Sims, J., Cummings, R.D., McEver, R.P., and Carlyon, J.A. (2006) Characterization of a sialic acid- and P-selectin glycoprotein ligand-1-independent adhesin activity in the granulocytotropic bacterium Anaplasma phagocytophilum. *Cell Microbiol* **8**: 1972–1984.
Rikihisa, Y. (2010) *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*: subversive manipulators of host cells. *Nature Publishing Group* **8**: 328–339.
Römling, U. (2009) Cyclic Di-GMP (c-Di-GMP) goes into host cells--c-Di-GMP signaling in the obligate intracellular pathogen *Anaplasma phagocytophilum. Journal of Bacteriology* **191**: 683–686.
Sarkar, A., Hellberg, L., Bhattacharyya, A., Behnen, M., Wang, K., Lord, J.M., et al. (2012) Infection with *Anaplasma phagocytophilum* activates the phosphatidylinositol 3-Kinase/Akt and NF-κB survival pathways in neutrophil granulocytes. *Infect. Immun.* **80**: 1615–1623.
Schaff, U.Y., Trott, K.A., Chase, S., Tam, K., Johns, J.L., Carlyon, J.A., et al. (2010) Neutrophils exposed to *A. phagocytophilum* under shear stress fail to fully activate, polarize, and transmigrate across inflamed endothelium. *Am. J. Physiol., Cell Physiol.* **299**: C87–96.
Sukumaran, B., Mastronunzio, J.E., Narasimhan, S., Fankhauser, S., Uchil, P.D., Levy, R., et al. (2011) *Anaplasma phagocytophilum* AptA modulates Erk1/2 signalling. *Cell Microbiol* **13**: 47–61.
Sutten, E.L., Norimine, J., Beare, P.A., Heinzen, R.A., Lopez, J.E., Morse, K., et al. (2010) *Anaplasma marginale* type IV secretion system proteins VirB2, VirB7, VirB11, and VirD4 are immunogenic components of a protective bacterial membrane vaccine. *Infect. Immun.* **78**: 1314–1325.
Thirumalapura, N.R., Qin, X., Kuriakose, J.A., and Walker, D.H. (2014) Complete Genome Sequence of *Ehrlichia muris* Strain AS145^{T}, a Model Monocytotropic *Ehrlichia* Strain. *Genome Announcements* **2**: e01234–13–e01234–13.
Thomas, R.J., Dumler, J.S., and Carlyon, J.A. (2009) Current management of human granulocytic anaplasmosis, human monocytic ehrlichiosis and *Ehrlichia ewingii* ehrlichiosis. *Expert Rev Anti Infect Ther* **7**: 709–722.

Thomas, S., Popov, V.L., and Walker, D.H. (2010) Exit mechanisms of the intracellular bacterium *Ehrlichia*. *PLoS ONE* **5**: e15775.
Thomas, V. and Fikrig, E. (2007) *Anaplasma phagocytophilum* specifically induces tyrosine phosphorylation of ROCK1 during infection. *Cell Microbiol* **9**: 1730–1737.
Troese, M.J. and Carlyon, J.A. (2009) *Anaplasma phagocytophilum* dense-cored organisms mediate cellular adherence through recognition of human P-selectin glycoprotein ligand 1. *Infect. Immun.* **77**: 4018–4027.
Wakeel, A., Dulk-Ras, den, A., Hooykaas, P.J.J., and McBride, J.W. (2011) *Ehrlichia chaffeensis* tandem repeat proteins and Ank200 are type 1 secretion system substrates related to the repeats-in-toxin exoprotein family. *Front Cell Infect Microbiol* **1**: 22.
Wakeel, A., Kuriakose, J.A., and McBride, J.W. (2009) An *Ehrlichia chaffeensis* tandem repeat protein interacts with multiple host targets involved in cell signaling, transcriptional regulation, and vesicle trafficking. *Infect. Immun.* **77**: 1734–1745.
Wang, X., Cheng, Z., Zhang, C., Kikuchi, T., and Rikihisa, Y. (2007) *Anaplasma phagocytophilum* p44 mRNA expression is differentially regulated in mammalian and tick host cells: involvement of the DNA binding protein ApxR. *Journal of Bacteriology* **189**: 8651–8659.
Xiong, Q. and Rikihisa, Y. (2012) Subversion of NPC1 pathway of cholesterol transport by *Anaplasma phagocytophilum*. *Cell Microbiol* **14**: 560–576.
Xiong, Q. and Rikihisa, Y. (2011) The prenylation inhibitor manumycin A reduces the viability of *Anaplasma phagocytophilum*. *J. Med. Microbiol.* **60**: 744–749.
Xiong, Q., Bao, W., Ge, Y., and Rikihisa, Y. (2008) *Ehrlichia ewingii* infection delays spontaneous neutrophil apoptosis through stabilization of mitochondria. *J. Infect. Dis.* **197**: 1110–1118.
Xiong, Q., Lin, M., and Rikihisa, Y. (2009) Cholesterol-dependent *Anaplasma phagocytophilum* exploits the low-density lipoprotein uptake pathway. *PLoS Pathog* **5**: e1000329.
Zhang, J.-Z., Popov, V.L., Gao, S., Walker, D.H., and Yu, X.-J. (2007) The developmental cycle of *Ehrlichia chaffeensis* in vertebrate cells. *Cell Microbiol* **9**: 610–618.
Zhang, J.-Z., Sinha, M., Luxon, B.A., and Yu, X.-J. (2004) Survival strategy of obligately intracellular *Ehrlichia chaffeensis*: novel modulation of immune response and host cell cycles. *Infect. Immun.* **72**: 498–507.
Zhu, B., Nethery, K.A., Kuriakose, J.A., Wakeel, A., Zhang, X., and McBride, J.W. (2009) Nuclear translocated *Ehrlichia chaffeensis* ankyrin protein interacts with a specific adenine-rich motif of host promoter and intronic *Alu* elements. *Infect. Immun.* **77**: 4243–4255.

RESULTATS

Partie 1

Identification des protéines de la membrane externe d'*E. ruminantium*

1. Préambule

E. ruminantium est une bactérie intracellulaire obligatoire qui possède un tropisme pour les cellules endothéliales et les neutrophiles. Son cycle de développement comporte deux phases: un stade ou la bactérie est sous formes de corps réticulé intracellulaire non-infectieux et métaboliquement actif et un stade où la bactérie se présente sous forme de corps élémentaire qui joue un rôle fondamental dans l'adhésion et l'infection des cellules hôtes (Jongejan et al. 1991). Chez les *Ehrlichia,* l'adhésion des bactéries est réalisée via des récepteurs de la membrane externe (Mohan Kumar et al. 2013). L'entrée et la prolifération impliquent la formation de cavéoles, petites invaginations de la membrane plasmique de l'hôte et qui mènent à l'incorporation des lipoprotéines et du cholestérol dans la membrane bactérienne (Lin and Rikihisa 2003). Après internalisation par la cellule hôte, les bactéries du genre *Ehrlichia* possèdent la capacité de se soustraire à cet environnement hostile en résidant dans un compartiment vacuolaire de type endosome qui ne fusionne pas avec les lysosomes (Barnewall, Rikihisa, and Lee 1997). En tant que bactérie intracellulaire stricte, *E. ruminantium* nécessite d'échanger des métabolites et des nutriments avec le cytoplasme de la cellule hôte. Ces évènements ne peuvent se faire que via des protéines de la membrane externe de la bactérie. De plus, des études sur les bactéries de la famille des *Anaplasmataceae* ont montré un rôle important des protéines de la membrane externe dans la stimulation de la réponse immune de l'hôte et dans la protection de celui-ci vis-à-vis d'une infection (Lopez et al. 2007; Ohashi et al. 1998). Depuis de nombreuses années, des vaccins expérimentaux sont testés comme moyen de lutte (vaccins inactivés, atténués) conférant des résultats peu ou pas efficaces (Pretorius et al. 2010). De plus, sur le terrain où plusieurs souches coexistent, la mise en place d'un vaccin conférant une immunité croisée pour une grande variété de souches est difficile à cause de la variabilité génétique qui existe entre les souches (Adakal et al. 2010). Ainsi, la mise en place d'un vaccin efficace contre toutes les souches nécessite une meilleure compréhension du pathogène et de ses mécanismes de virulence notamment par l'identification des protéines de la membrane externe.

Afin d'identifier les protéines de la membrane externe *d'E. ruminantium*, nous avons optimisé un protocole de purification de la membrane externe et utilisé une approche protéomique (nanoLC-MS/MS). Cette analyse, couplée à une analyse bioinformatique, a permis d'identifier 46 protéines uniques dans la fraction membranaire externe dont 18

protéines de la membrane externe. Cette analyse confirme la présence de protéines membranaires connues impliquées dans la structure et la biogénèse de la membrane externe, dans le transport et la virulence. Nous avons aussi identifié des protéines de la membrane externe de fonctions inconnues. Ce travail représente à ce jour la caractérisation la plus complète de la fraction de la membrane externe des *Ehrlichia*. Cette étude indique que des expériences de fractionnement subcellulaire adaptées combinées à des approches protéomiques et bioinformatiques sont un outil puissant pour déterminer la localisation subcellulaire des protéines chez les *Ehrlichia*. Les nouvelles protéines hypothétiques identifiées représentent de bons candidats vaccinaux.

Ainsi, les protéines de la membrane externe sont une source riche de candidats pour le développement de nouveaux traitements nécessaires contre l'émergence des maladies infectieuses et des bactéries multi-résistantes. Ces résultats permettent d'avoir une meilleure vision de l'architecture de la membrane externe d'*Ehrlichia* et ouvrent la voie au développement de nouvelles stratégies thérapeutiques pouvant perturber le fonctionnement de la membrane externe de la bactérie.

2. Publication: Proteomic profiling of the outer membrane fraction of the obligate intracellular bacterial pathogen *Ehrlichia ruminantium*

Proteomic profiling of the outer membrane fraction of the obligate intracellular bacterial pathogen *Ehrlichia ruminantium*

Amal Moumène[a,b,c], Isabel Marcelino[d,e], Miguel Ventosa[d,e], Olivier Gros[f], Thierry Lefrançois[b], Nathalie Vachiéry[a,b], Damien F. Meyer[a,b] * and Ana V. Coelho[e] *

[a] CIRAD, UMR CMAEE, Site de Duclos, Prise d'eau, F-97170 Petit-Bourg, Guadeloupe, France

[b] INRA, UMR1309 CMAEE, F-34398 Montpellier, France

[c] Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

[d] Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2780-901 Oeiras, Portugal

[e] Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal

[f] Université des Antilles et de la Guyane, Institut de Biologie Paris Seine, UMR7138 UPMC-CNRS, Equipe Biologie de la Mangrove, UFR des Sciences Exactes et Naturelles, Département de Biologie, BP 592, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

* DFM and AVC contributed equally to the research and are joint senior authors

Corresponding author: damien.meyer@cirad.fr

Title

Proteomic profiling of the outer membrane fraction of the obligate intracellular bacterial pathogen *Ehrlichia ruminantium*

Amal Moumène[a,b,c], Isabel Marcelino[d,e], Miguel Ventosa[d,e], Olivier Gros[f], Thierry Lefrançois[b], Nathalie Vachiéry[a,b], Damien F. Meyer[a,b] * and Ana V. Coelho[e] *

[a] CIRAD, UMR CMAEE, Site de Duclos, Prise d'eau, F-97170 Petit-Bourg, Guadeloupe, France

[b] INRA, UMR1309 CMAEE, F-34398 Montpellier, France

[c] Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

[d] Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2780-901 Oeiras, Portugal

[e] Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal

[f] Université des Antilles et de la Guyane, Institut de Biologie Paris Seine, UMR7138 UPMC-CNRS, Equipe Biologie de la Mangrove, UFR des Sciences Exactes et Naturelles, Département de Biologie, BP 592, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

* DFM and AVC contributed equally to the research and are joint senior authors

Corresponding author: damien.meyer@cirad.fr

Running title: Outer membrane proteome of *Ehrlichia ruminantium*

Abstract

The outer membrane proteins (OMPs) of Gram-negative bacteria play a crucial role in virulence and pathogenesis. Identification of these proteins represents an important goal for bacterial proteomics, because it aids in vaccine development. Here, we have developed such an approach for *Ehrlichia ruminantium*, the obligate intracellular bacterium that causes heartwater. A preliminary whole proteome analysis of elementary bodies, the extracellular infectious form of the bacterium, had been performed previously, but information is limited about OMPs in this organism and about their role in the protective immune response. Identification of OMPs is also essential for understanding *Ehrlichia*'s OM architecture, and how the bacterium interacts with the host cell environment. First, we developed an OMP extraction method using the ionic detergent sarkosyl, which enriched the OM fraction. Second, proteins were separated via one-dimensional electrophoresis, and digested peptides were analyzed via nano-liquid chromatographic separation coupled with mass spectrometry (LC-MALDI-TOF/TOF). Of 46 unique proteins identified in the OM fraction, 18 (39%) were OMPs, including 8 proteins involved in cell structure and biogenesis, 4 in transport/virulence, 1 porin, and 5 proteins of unknown function. These experimental data were compared to the predicted subcellular localization of the entire *E. ruminantium* proteome, using three different algorithms. This work represents the most complete proteome characterization of the OM fraction in *Ehrlichia spp*. The study indicates that suitable subcellular fractionation experiments combined with straightforward computational analysis approaches are powerful for determining the predominant subcellular localization of the experimentally observed proteins. We identified proteins potentially involved in *E. ruminantium* pathogenesis,

which are good novel targets for candidate vaccines. Thus, combining bioinformatics and proteomics, we discovered new OMPs for *E. ruminantium* that are valuable data for those investigating new vaccines against this organism. In summary, we provide both pioneering data and novel insights into the pathogenesis of this obligate intracellular bacterium.

Importance

Ehrlichiae are obligate intracellular bacteria with a unique developmental cycle that includes attaching to and entering eukaryotic host cells, a process mediated by proteins in their outer membrane (OM). Thus far, few experimental data on ehrlichial OM proteins are available. To gain insight into the protein composition of the ehrlichial OM, we performed proteome analysis on OM fractions from *Ehrlichia ruminantium* elementary bodies, the infectious form of this bacterium. We compared our experimental results with an *in silico* analysis of the *E. ruminantium* proteome. We identified 18 proteins, whose OM localization was supported by both studies, and were, therefore, very likely to be located in the *E. ruminantium* OM. Among these proteins, 6 are completely new discovered OMPs and are therefore of importance as potential vaccine antigens. These results provide the first comprehensive overview of OM proteins in an *Ehrlichia* species and pave the way for developing novel therapeutic strategies to disrupt the OM or processes essential for its function.

Introduction

The *Rickettsiales Ehrlichia ruminantium* is an obligate intracellular bacterium that causes heartwater, a fatal tick-borne disease of ruminants, which is found in the islands of the Indian Ocean and the Caribbean, and in Africa (1). *E. ruminantium* is transmitted by *Amblyomma* ticks and infects the endothelium of blood vessels. It has a complex life cycle with two distinct developmental forms found within mammalian host cells (2). Initially, the infectious forms of the bacterium (elementary bodies, or EBs) adhere to host target cells and are internalized. Then, inside of intracytoplasmic vacuoles, they differentiate into a replicative, non-infectious form, the reticulate body (RB). After 5 to 6 days of intracellular multiplication, disruption of host cells leads to the release of numerous infectious EBs, initiating a new infectious cycle (1, 3).

Current control methods for heartwater consist of a combination of vector control, using acaricides, and immunization against *E. ruminantium*. Different types of vaccines (inactivated, attenuated, recombinant) are currently being tested experimentally, but they have displayed limited efficacy, thus far, due to the genetic and antigenic diversity of *E. ruminantium* strains (3-8). At this time, the only commercially available vaccine is based on the administration of infected blood to ruminants, followed by treatment with antibiotics; however, this remains an expensive, high-risk method (3).

Many studies of Gram-negative bacteria, such as *Legionella pneumophila*, *Bartonella henselae*, *Pseudomonas syringae*, *Campylobacter jejuni*, and *Mannheimia haemolytica*, have focused on outer membrane proteins (OMPs), because they have proven to be good targets for vaccine development (9-13). Indeed, the OM of such pathogens represents an important dynamic interface between the bacterium and its

environment. It serves as a selective barrier controlling the passage of nutrients and waste products into and out of the cell, and it also creates a chemically distinct periplasmic compartment, where important processes, such as the degradation of harmful substances from the environment or certain types of respiration, can occur (14, 15). OMPs are involved in the integrity and stability of the bacterial envelope, passive and active transport of substrates and nutrients, cell-to-cell communication, adhesion to host cells, and virulence (16).

Prospective proteomic analysis of *E. ruminantium*, cultivated in host endothelial cells, has already provided information about OMPs that are potentially implicated in bacterial infection and survival, such as members of the major antigenic protein (*map)* gene cluster (17, 18). Despite significant evidence implicating this gene family in immune protection in *Ehrlichia* and *Anaplasma* (19, 20) and even strain penetrance in *Anaplasma* (21), our understanding of the biological role of this gene family is incomplete. However, studies on the differential expression of genes encoding OMPs has permitted us to understand the adaptation of these bacteria to the environment inside their vector, the tick, and to transmission to the mammalian host (22, 23).

The aim of this study was to characterize the proteome of the OM fraction from infectious *E. ruminantium* EBs. To obtain an enriched OM fraction, we optimized a sarkosyl-based enrichment protocol that selectively solubilizes the inner and cytoplasmic membranes of Gram-negative bacteria, with no effect on the OM subcellular fraction (24). We identified 46 unique proteins in the OM fraction using one-dimensional gel electrophoresis coupled with liquid chromatography-mass spectrometry (1DE-nanoLC-MALDI-TOF/TOF). Of these, 18 were known or predicted prototypical OMPs, while the others were of inner membrane (n=5) or cytoplasmic

(n=23) origin or were chaperones. We compared our experimental results to the total set of *E. ruminantium* OMPs by combining results from three subcellular localization prediction algorithms and 34% of the total OMPs predicted from the genome were detected in the obtained OM fraction. We concluded that our method enriched OMPs. These results provide a better understanding of *Ehrlichia* OM architecture and may lead to the identification of potential vaccine candidates.

Materials and Methods

Ehrlichia ruminantium cultivation

E. ruminantium strain Gardel (from Guadeloupe, FWI) was routinely propagated in bovine aorta endothelial cells (BAE) as previously described (25). One-hundred and twenty hours post-infection, when cell lysis occurs, infectious EBs were harvested and purified using a multistep, 20,000 × *g* centrifugation protocol, as described elsewhere (26, 27). Purified EBs were stored at -80°C in sucrose-phosphate-glutamate (SPG) buffer, pH 7.4.

Preparation of the OM fraction from *E. ruminantium* EBs

Subcellular fractionation was performed as described by Ohashi *et al.* (28), modified as follows. Purified EBs stored in SPG were washed in phosphate-buffered saline (PBS, pH 7.4) with a protease inhibitor cocktail (Roche), at 20,000 × *g* for 30 min at 4°C. Protein content was measured with the microBCA quantification kit (Sigma), according to the manufacturer's instructions. Five hundred micrograms EBs were pelleted and resuspended in PBS containing 0.1% (v:v) sodium N-laurosyl sarcosine (sarkosyl; Sigma), DNAse (50 µg/mL), RNAse (50 µg/mL), $MgCl_2$ (2.5 mM), and protease inhibitors (Roche), and then incubated for 30 min at 37°C. The sarkosyl

treatment was repeated twice, followed by ultracentrifugation at 20,000 × *g* for 30 min at 4°C (Figure 1). After the first separation, the insoluble pellet containing the OM fraction was washed twice in PBS and centrifuged at 20,000 × *g* for 30 min at 4°C to remove residual detergent (Step 2); the final pellet was resuspended in PBS containing protease inhibitors, and then stored at 4°C. Total protein concentration was determined using the 2D Quant Kit™ (GE Healthcare). Independent biological triplicates were carried out for OMP characterization (Figure 1).

Evaluation of OM enrichment protocol

1. Transmission Electron Microscopy (TEM)

Samples were pre-fixed at 4°C in 2.5 % (v/v) glutaraldehyde in PBS (pH 7.2). After a brief rinse with 1 × PBS, samples (intact EBs or OM complex) were fixed for 45 min at 25°C in 1% (w/v) osmium tetroxide in the same buffer, rinsed in distilled water and post-fixed with 2% (w/v) aqueous uranyl acetate for 1 h at 25°C before being embedded in epoxy resin. Two grids containing 4–5 ultrathin sections (60 nm thick) were observed using a Tecnai G2 TEM at 200 kV (29). The TEM micrographs presented in this study are representative of all samples.

2. SDS-PAGE and Western blots to monitor OM fraction

Biological samples (15 µg) were precipitated in acetone for 3 h at -20°C and centrifuged at 20,000 × *g* for 10 min at 4°C. The pellet was solubilized in NuPAGE® LDS Sample Buffer loaded on NuPAGE® Novex® 4–12% Bis-Tris polyacrylamide gels, and electrophoresis was carried out for 40 min at 200 V. Proteins were transferred to polyvinylidene difluoride (PVDF) membranes (Millipore, USA). The membranes were blocked for 1 h in PBS with 0.05% (v/v) Tween 20 and 5% (w/v)

milk, and then incubated with anti-MAP1 mouse monoclonal antibody (mAB) (4F10B4, Abcam) at a dilution of 1:2,000 for 1 h. Anti-Map1 monoclonal antibody was used as a specific OM marker. Membranes were washed three times in PBS with 0.05% (v/v) Tween 20 for 10 min, followed by incubation with the appropriate phosphatase alkaline-conjugated secondary antibodies (Sigma) at a 1:2,000 dilution for 1 h. Finally, membranes were developed using 5-bromo-4-chloro-3'-indolyphosphate/nitro-blue tetrazolium (BCIP/NBT) substrate (Roche) (17).

Proteome Characterization

1. 1D gel electrophoresis for proteomics analysis

Forty μg intact EBs or OM fraction (from ERGp45, p52, and p57) were precipitated in acetone for 3 h at -20°C and centrifuged at 20,000 × *g* for 10 min at 4°C. Pellets were resuspended in 5 μL solubilization buffer [7 M urea, 2 M thiourea, 4% (w/v) 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), and 30 mM Tris; Step 3 in Figure 1]. After protein solubilization, 6 μL loading buffer [0.5 M dithiothreitol (DTT), 10% (w/v) sodium dodecyl sulfate (SDS), 250 mM Tris, 30% (v/v) glycerol, and 0.02% (w/v) bromophenol blue] was added. Samples were vortexed, and 9 μL water was added followed by agitation overnight at room temperature. Finally, samples were centrifuged at 16,000 × *g* for 2 min, and supernatants were loaded on NuPAGE® Novex® 4–12% Bis-Tris polycacrylamide gels; electrophoresis was performed for 40 min at 200 V. Gels were stained for 24 h using colloidal Coomasie Blue, and then washed 3 times in double distilled water (17).

2. In-gel digestion

For the evaluation of the optimized protocol to obtain an OMP enriched fraction, the more intense gel bands were excised. Previously to the NanoLC-MALDI-TOF/TOF

analysis and in order to extend the number of proteins identified starting from simpler peptide digests, the OMP enriched fraction was separated by SDS-PAGE and each gel lanes was sliced. For in-gel digestion each band or slice was cut into 1 mm^3 gel pieces, and Coomasie Blue was washed off with alternating water and 50% (v/v) acetonitrile (ACN) treatments until the gel pieces were transparent. Proteins were in-gel reduced with 10 mM dithiothreitol (DTT), alkylated with 55 mM iodoacetamide. Next, 6.7 ng/μL modified porcine trypsin (Promega) in 50 mM NH_4CO_3 was added to each gel band/slice. Digestion was performed at 37°C overnight. Peptides were extracted from the gel by washing it with 5% (v/v) formic acid, followed by two ACN washes. Digestion supernatants and extracted peptides were added, dried in a SpeedVac concentrator, and reconstituted in 5% (v/v) formic acid (30).

3. NanoLC-MALDI-TOF/TOF analysis

Chromatographic peptide separation was performed on a Thermo EASY-nLC 1000 with a pre-column Acclaim PepMap 100 C18 (75 μm × 2 cm) used as the Peptrap and an Acclaim PepMap RSLC C18 (50 μm × 15 cm) as the chromatographic separation column (Step 4, Figure 1). A chromatographic gradient was established using mixed volumes of 0.1% (v/v) formic acid in water (buffer A) and 0.1% (v/v) formic acid in acetonitrile (buffer B, all LC-MS grade, from MERCK); peptides were eluted at a constant rate of 2 mL/min for 40 min in 5–40% (v/v) buffer A, according to their hydrophilic/hydrophobic properties. Peptide fractions were spotted onto MALDI plates and co-crystalized with 5 mg/mL alpha-cyano-4-hydroxycinnamic acid using a Micro-Spotter (Sunchrom). Peptide mass spectra were acquired with an Applied Biosystems 4800 Plus MALDI TOF/TOF Analyzer apparatus in both MS and MS/MS mode. Positively charged ions were analyzed in the reflectron mode over an

m/z range of 800–3,500 Da. Each MS spectrum was obtained in result-independent acquisition mode with a total of 800 laser shots per spectra and a fixed laser intensity of 3,500 V. Calibration was performed using Des-Arg-bradykinin (904.468 Da), angiotensin 1 (1,296.685 Da), Glu-Fibrinopeptide B (1,570.677 Da), ACTH (1–17 clip) (2,093.087 Da), and ACTH (18–39 clip) (2,465.199 Da) (Calibration Mix from Applied Biosystems). Fifteen s/n best precursors from each MS spectrum were selected for MS/MS analysis. MS/MS analyses were performed using collision-induced dissociation (CID) assisted with air, using a collision energy of 1 kV and a gas pressure of 10^6 Torr. Two thousand laser shots were collected for each MS/MS spectrum using a fixed laser intensity of 4,500 V. Raw data were generated using 4000 Series Explorer Software v3.0 RC1 (Applied Biosystems, Foster City, CA, USA), and all contaminant m/z peaks originating from human keratin, trypsin autodigestion, or matrix were placed on the exclusion list used to generate the peptide mass list used in the database search (17).

4. Database query

To identify proteins, Mascot generic format files combining MS and MS/MS spectra were used to interrogate a non-redundant protein database using a local Mascot v2.2 license from Matrix Science and the Global Protein Server (GPS) v3.6 (Applied Biosystems). Search parameters for the MS/MS spectra were as follows: i) the Uniprot (2013) sequence database (*E. ruminantium* with isoforms) was used; ii) taxonomy was set to “all entries” (302,409); iii) variable modifications were considered [i.e., carbamidomethylation (Cys), deamidation (Asn and Gln), and oxidation (Met, Pro, Lys, Arg)]; iv) two missed cleavage sites were allowed; v) precursor tolerance was set to 50 ppm and MS/MS fragment tolerance to 0.5 Da; vi)

peptide charge was 1+; and vii) the algorithm used trypsin as the enzyme. A protein candidate provided by this MS/MS search was considered valid if the global Mascot score was >40 at a significance level of $p<0.05$, if at least one peptide was identified with 95% confidence, and if it was found in at least two of the three biological replicates.

In silico genome analysis

The publicly available proteome of the *E. ruminantium* strain Gardel, which was extracted from the Uniprot database (31) in FASTA format, was used for bioinformatics studies. The subcellular localization of the 948 *E. ruminantium* protein-coding genes was predicted using three global programs: PSORTb 3.0 (32), CELLO 2.5 (33), and MetaLocGramN (34). The predicted utilization locations of each protein were filtered from raw software output using in-house scripts written in the R programming language and exported to Excel. In some cases, CELLO 2.5 predicted multiple localization sites for the same protein. The proteins involved were grouped under the heading “unknown localization.”

As a result of the varying predictions for a given protein, the consensus prediction was calculated using a majority vote procedure. If two of three algorithms agreed on localization, this localization was attributed to the protein. As for the remaining results, when outer or inner membrane localization was predicted by only one program, protein subcellular localization was refined manually, based on the experimental data in the literature, or the presence of signal peptides, transmembrane domains using dedicated algorithms (Table 1; Excel file in Supplementary data).

Results

Enrichment of *E. ruminantium* OM fraction

The first step in this study was to recover most of the OM complex with minimal contamination by cytoplasmic and inner membrane fractions. To do this, we used sarkosyl, an ionic detergent commonly used in the purification of OMs in Gram-negative bacteria, because it selectively solubilizes cytoplasmic and inner membranes while conserving the integrity of the OM (24). Figure 1 shows the workflow used to obtain the OM fraction. To assess protocol efficacy, samples were harvested at critical time points during the purification process, and their quality was evaluated using TEM, SDS-PAGE to identify proteins in the most intense bands, and Western blotting (Figure 2). After sarkosyl treatment of intact EBs (Figure 2A), empty shells with spherical morphology, corresponding to the OM fraction, were observed (Figure 2B). These OM complexes, with a diameter of approximately 200 nm, appeared to be devoid of inner membrane and cytoplasm components, in contrast to intact EBs (Figure 2A). Comparative protein migration profiles of the different fractions (intact EBs, E; sarkosyl soluble fractions, S; and outer membrane fractions, OMs) were analyzed using SDS-PAGE (Figure 2C), and each subcellular fraction displayed a distinct migration pattern. The OM preparation showed prominent bands at approximately 134, 63, 55, 41, 37, and 29 kDa. The most abundant proteins, in the 30 kDa range, may represent Map1 protein family. When the different fractions were analyzed via Western blot using a monoclonal antibody against Map1 (a specific OM marker), intact EBs (the positive control) displayed a strong ~30 kDa band corresponding to Map1 (Figure 2D). This protein was detected in the OM fraction but not in the soluble fraction, confirming the efficacy of the purification protocol (Figure 2D). Altogether, these results clearly indicate that the insoluble sarkosyl fraction was

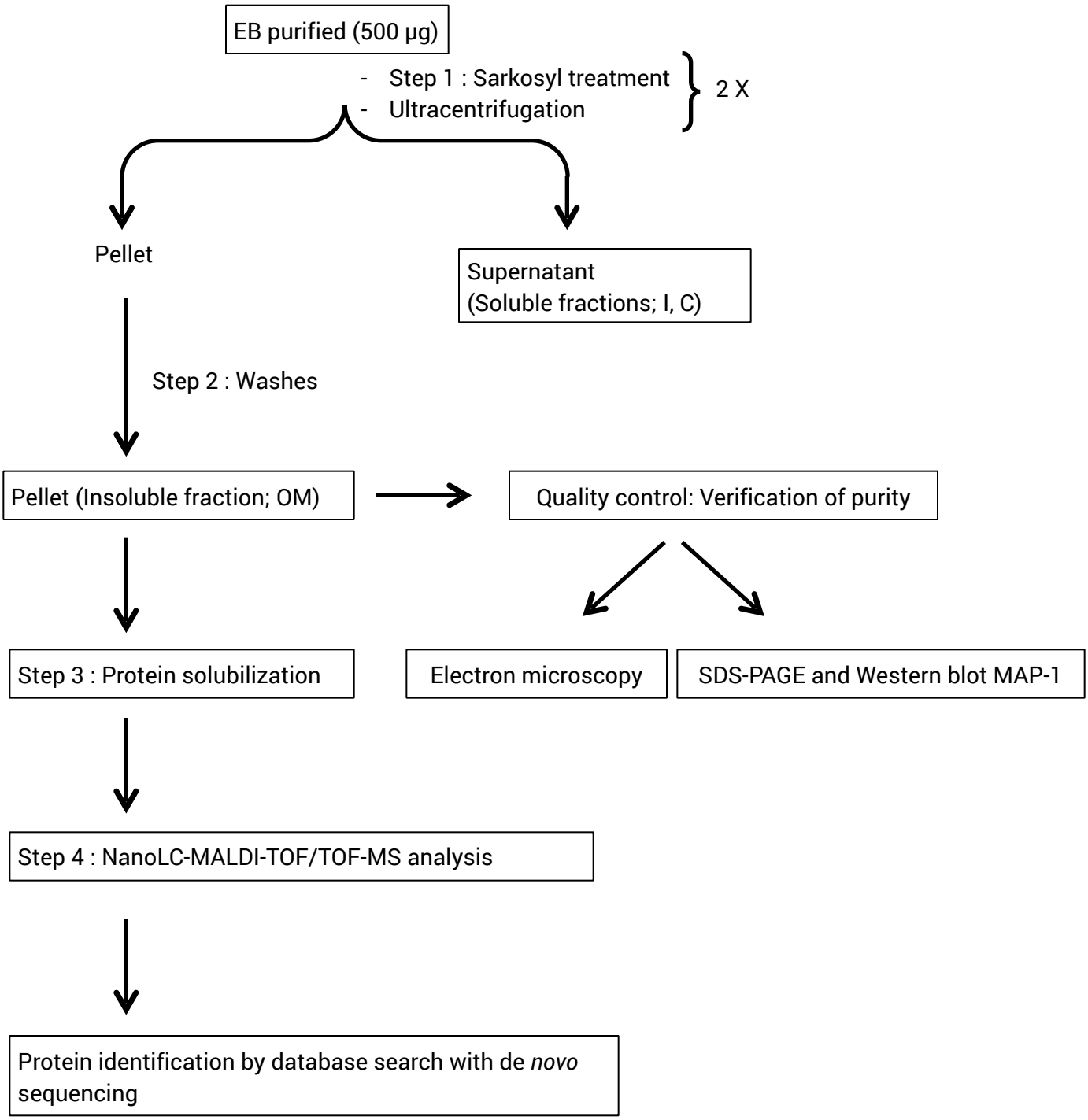

EB purified (500 µg)
- Step 1 : Sarkosyl treatment
- Ultracentrifugation
2 X
Pellet
Supernatant
(Soluble fractions; I, C)
Step 2 : Washes
Pellet (Insoluble fraction; OM)
Quality control: Verification of purity
Step 3 : Protein solubilization
Electron microscopy
SDS-PAGE and Western blot MAP-1
Step 4 : NanoLC-MALDI-TOF/TOF-MS analysis
Protein identification by database search with de *novo* sequencing

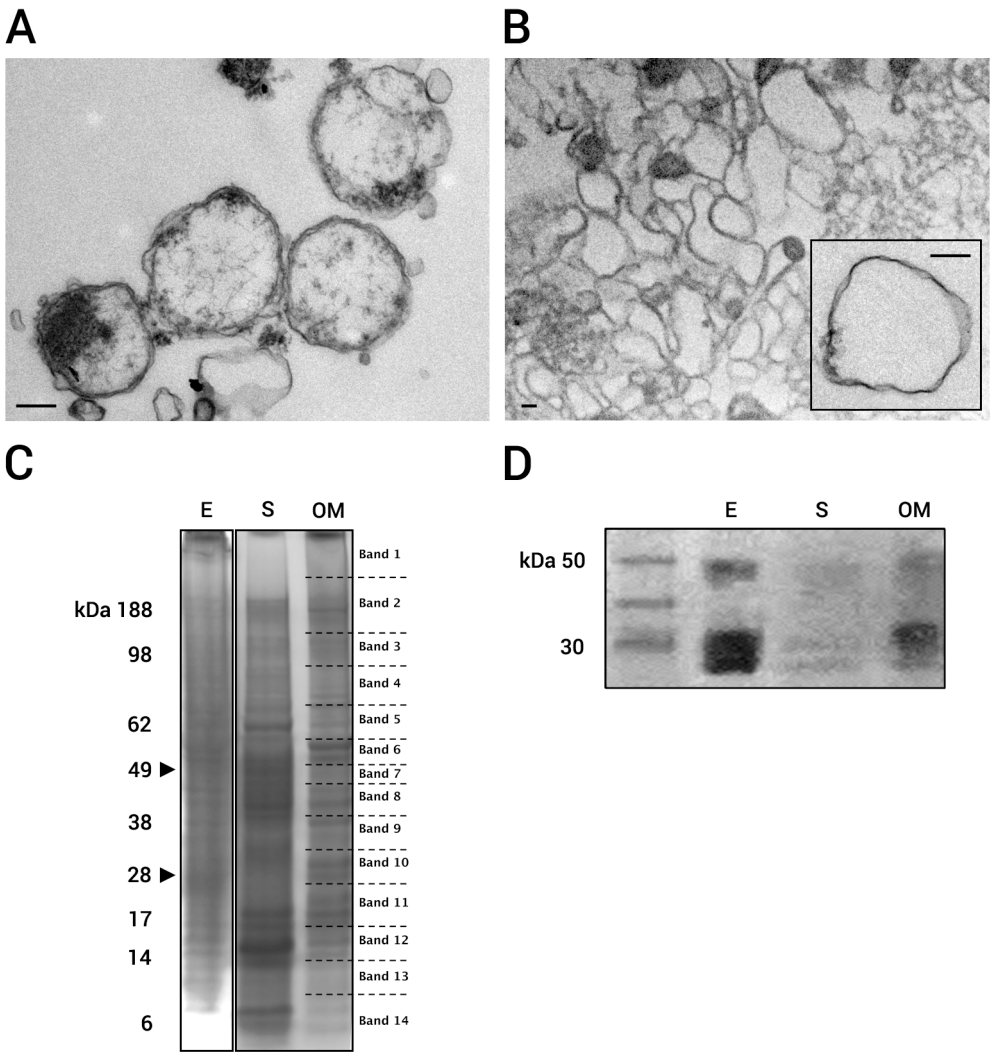
Figure 1

strongly enriched with *E. ruminantium* OM complexes.

In silico subcellular localization prediction of *E. ruminantium* proteins

We utilized a combination of three computational prediction tools, CELLO 2.5, PSORTb 3.0, and MetaLocGramN, to predict subcellular localization in the entire *E. ruminantium* proteome. These programs have been used to identify OMPs in several Gram-negative bacterial species (36-38). Though the programs made diverse subcellular localization predictions for the same proteins, the combination of different predictors minimizes the risk of false positives for OMP prediction. PSORTb 3.0, CELLO 2.5, and MetaLocGramN predicted 490, 461, and 526 cytoplasmic proteins in *E. ruminantium* (~50 % of total proteins), respectively (Table 1). CELLO 2.5 predicted 11.5% of proteins were inner membrane proteins (IMPs), whereas the two other programs predicted roughly twice as many (20%). CELLO 2.5 identified the highest proportion of OMPs (9.4%, 90/948), followed by MetaLocGramN (7.4%, 71/948) and PSORTb 3.0 (1.1%, 11/948). PSORTb 3.0 could not predict the localization of 236 proteins, while CELLO could not provide predictions for 256.

Altogether, we predicted that the total proteome of *E. ruminantium* (948 proteins) consisted of 53% (499/948) cytoplasmic proteins, 13% (124/948) IMPs, and 5.4% (52/948) OMPs (Table 1). In Figure 3, the number of proteins in each Venn diagram compartment corresponds the consensus prediction correctly predicted by an algorithm for a given subcellular localization. Of the 52 OMPs identified using consensus predictions, 6 were identified by all three programs. Twenty-one were predicted by only a single program: 19 for CELLO 2.5 and 2 for MetaLocGramN. CELLO 2.5 predicted the highest number of consensus OMPs (50), followed by MetaLocGramN (33) and PSORTb 3.0 (6). All three programs identified two hundred

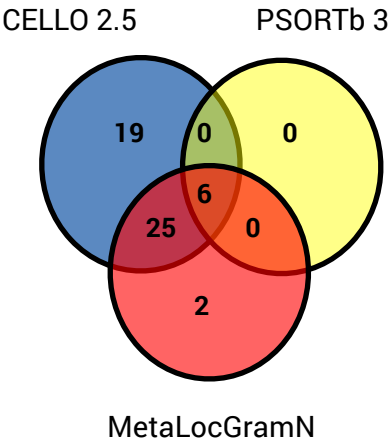

A
B
C
E
S
OM
kDa 188
98
62
49
38
28
17
14
6
Band 1
Band 2
Band 3
Band 4
Band 5
Band 6
Band 7
Band 8
Band 9
Band 10
Band 11
Band 12
Band 13
Band 14
D
E
S
OM
kDa 50
30

and ninety cytoplasmic proteins. CELLO 2.5 predicted the highest number of cytoplasmic proteins, whereas PSORTb 3.0 predicted the lowest.

Identification of proteins in the *E. ruminantium* OM fraction

OM fractions prepared from three biological replicates were analyzed individually using 1DE-nanoLC-MALDI-TOF/TOF MS. The proteins identified are presented in Table 3. Of the 46 non-redundant proteins identified in the OM fraction, 41 had known functions (either characterized experimentally or annotated via high sequence similarity), and the remaining five proteins were classified as hypothetical proteins. Several of these proteins (e.g. ERGA_CDS_04510, ERGA_CDS_04580) are conserved among members of *Anaplasmataceae*. Of the 46 proteins identified, 39% were indeed OMPs (18/46), 11% were IMPs (5/46), and 50% (23/46) were cytoplasmic. These proteins were classified into four functional groups: structural and transport proteins, biogenesis proteins (e.g. BamA, ERGA_CDS_08660), virulence proteins, and proteins involved in metabolic processes (e.g. GroEL, ERGA_CDS_06640 and Ef-Tu, ERGA_CDS_01580). Several ribosomal proteins and chaperones were also identified. Of the 18 OMPs identified, 5 belonged to the well-known MAP1 family (Map1, Map1+1, Map1-6, Map1-13, and Map1-14), 2 comprised β-barrel assembly machinery (BamA and BamD), 3 were components of the type IV secretion system (VirB9-1, VirB9-2, and VirB10), 1 was a porin, and 1 was a major ferric iron-binding protein. The six putative uncharacterized proteins had neither functional annotations in UniProt, nor hits in the Pfam database. Two of these (ERGA_CDS_04580, ERGA_CDS_05150) were predicted by SignalP to contain signal peptides. The first had no homology with known proteins and seemed to be unique in the *E. ruminantium* genome, whereas the second had similarity to

Outer membrane proteins



Inner membrane proteins

Cytoplasmic proteins

Figure 3

ECH_0525, an ortholog of Esp73, an OMP in *Anaplasma phagocytophilum*.

In summary, our study increased the number of OMPs experimentally identified accounting for 34% of total predicted OMPs in *E. ruminantium* (18/52), whereas the total number OMPs account only for 5.5% of *E. ruminantium* proteome (52/948). Thus, the OM purification process described enriched OMPs.

Discussion

The OM of Gram-negative bacteria is an important interface between the outside and inside of the cell. It protects bacteria against hostile environments. OMPs fulfill a number of crucial functions, such as supporting the biogenesis and integrity of the OM and acting as porins and virulence factors, playing a fundamental role in adherence to host cells, invasion, and evasion of host-defense mechanisms (39).

The purification of OMs is a key step in the identification of OMPs. Several methods, such as isopycnic centrifugation using a sucrose gradient, addition of Triton X-100, and carbonate extraction protocols, have been tested in bacteria (11-13, 40). However, the sarkosyl solubilization strategy, which solubilizes IM proteins and separates IM and OM proteins (24), has become the preferred method for many Gram-negative bacteria, due to the higher purity and better reproducibility of the OM extracts obtained in this manner (10, 41, 42). By applying this method to *E. ruminantium* EBs, we obtained a highly enriched OM fraction. Our proteomic analysis led to the identification of 18 unique OMPs corresponding to 34% of total cell OMPs. The low percentage of sarkosyl-insoluble proteins obtained may be due to excessive washing of the pellets after sarkosyl treatment, resulting in loss of proteins or lysis of cells (12, 25). In addition, OMP extraction was performed on the extracellular, infectious form of *Ehrlichia*. It is likely that only certain *E. ruminantium* proteins are

expressed at a given life cycle stage (43). For instance, expression of most *E. chaffeensis* proteins varies depending on host and vector environments and stage of development (44, 45).

We also analyzed the entire *E. ruminantium* proteome to determine the theoretical subcellular localization of all proteins (OM, IM, cytoplasmic, periplasmic, or extracellular). These *in silico* predictions allowed us to estimate the quality of the enrichment of OMPs in the OM fraction obtained using our purification protocol. PSORTb 3.0 is one of the most precise subcellular localization predictor for many Gram-negative bacteria (32). It uses a combination of factors based on motif and profile analyses, e.g. the presence of signal peptides, OM motifs, transmembrane helices, and similarity to proteins with known localization (32). However, in this study, it returned a high number of proteins with unknown localization (236 or 24.8% of total proteins). This problem may be due to the absence of significant sequence similarity between some *E. ruminantium* proteins and proteins in the PSORTb 3.0 database. Similar results have been observed in numerous other bacteria (34). Consequently, we chose two other computational localization predictors to overcome this weakness. CELLO 2.5 has the advantage of using multiple Support Vector Machines (SVMs) to analyze four types of protein descriptors, including amino acid composition, dipeptide composition, partitioned amino acid composition, and frequency of residues with particular physicochemical properties (33), yielding better predictive performance (33). However, in our study, CELLO 2.5 predicted multiple localization sites for 256 proteins that were subsequently grouped in a “unknown localization” category (36). Finally, we included MetaLocGramN program, a meta-predictor that combines multiple primary methods, including general subcellular localization, signal peptide

predictors, transmembrane helix predictors, and beta barrel OMP predictors (34). The combination of results from these three programs improved the accuracy of subcellular localization predictions (46, 13, 36).

Collectively, our bioinformatics analysis predicts that 5.4% of the annotated genes in the *E. ruminantium* genome are OMPs. Analyses of other Gram-negative bacteria have identified approximately the same percentage of predicted OMPs. For example, an analysis employing 10 different predictors to analyze the *Pasteurella multocida* genome identified 98 OMPs in an avian strain and 107 in a porcine strain (4.8% and 5.0% of total proteins, respectively) (47). Similarly, prediction of the subcellular localization of *P. syringae* Lz4W proteins, performed using PSORTb 3.0, revealed that 148 out of a total of 1,479 proteins (10%) were OMPs (11). In addition, we compared our results to those obtained experimentally from many other bacteria. In *L. pneumophila*, OM and surface-exposed proteome analyses using cellular fractionation and fluorescent labeling led to the identification of OMPs accounting for 8.5% of total proteins (9). These results suggest that our prediction of *E. ruminantium* OMPs yielded a reasonable identification rate.

We experimentally identified a total of 46 non-redundant proteins in the OM fraction, 18 of which were clearly classified as OMPs. These 18 OMPs correspond to 1.9% of the entire *E. ruminantium* proteome (18/948) and 34.6% of predicted OMPs in the entire proteome (18/52). Previous studies on the total *E. ruminantium* proteome have identified 64 non-redundant proteins including 8 OMPs (17). Thus, as expected, enriching the OM fraction resulted in an increased number of OMPs being identified. Some of these OMPs have known functions and include proteins of the Map1 cluster (48), BamA/D (49), VirB9-1 (50), VirB9-2, VirB10 (51), a porin (52), and major ferric

iron-binding protein (53). We also characterized five proteins classified as hypothetical but predicted to be OMPs, including ERGA_CDS_04510, 03960, 02510, 02370, and 05150. BLAST search on ERGA_CDS_05150 revealed an ortholog in *Ehrlichia chaffeensis*, Esp73; an ortholog to *A. phagocytophilum* Asp55 and Asp62, that is predicted to contain 22 transmembrane β-strands forming a β-barrel and, thus, may be involved in membrane transport (54). Further functional characterization of these newly discovered OMPs should be carried out to evaluate their potential as protective antigens.

Map1, the immunodominant, major OMP expressed by *E. ruminantium* in the mammalian host, is encoded by a member of a multigene family comprising 16 paralogs (55). The number of Map1 family proteins detected in this study (n=5: Map1, Map1+1, Map1-6, Map1-14, and Map1-13) was greater than that detected in a previous proteomic analysis (17). These proteins are known to be differentially transcribed *in vitro* in endothelial and tick cell cultures (55, 56) and are well conserved, since *omp-1*, *msp2*, *p44*, *p30*, and *map-1* belong to a superfamily harboring the PF01617 Pfam domain (1). Map1 family proteins are considered priority targets for candidate vaccines (57), as they are potentially involved in *E. ruminantium* adaptation to the mammalian host and its vector, the tick (18). However, few data are currently available on the expression and characterization of Map1 family proteins throughout the bacterial life cycle (17).

Proteins of the β-barrel Assembly Machinery (BAM) complex are involved in diverse cellular functions, including solute transport, protein secretion, and assembly of protein and lipid components of the OM (58). They account for the vast majority of bacterial OMPs and are essential for bacterial viability and function (59). The

insertion of proteins in the OM depends on a protein complex that contains the OMP BamA and four associated lipoproteins (BamB, C, D, and E) (60). BamA (ERGA_CDS_08660) and BamD (ERGA_CDS_08100) were identified in our experimental analysis. BamA proteins are essential for the biogenesis of β-barrel OMPs and play a central part in OMP assembly (61-63). It has been observed that reducing the levels of BamA significantly affects the ability of the β-barrel membrane protein OprF to localize to the OM, showing its essential role in OM biogenesis (63)., BamD is the only essential lipoprotein in the BAM complex (64), and it is highly conserved in Gram-negative bacteria as well (65).

Many bacterial species use specialized secretion systems to transfer macromolecules across membranes (66). The type IV secretion system (T4SS) translocates DNA or proteins across membranes directly into eukaryotic host cells to subvert host cellular functions. Consequently, the proteins that make up this system represent crucial bacterial virulence determinants in important human pathogens such as *B. henselae*, *Helicobacter pylori*, *L. pneumophila*, *Bordetella pertussis*, and *Brucella melitensis* (67, 68). In this study, we identified three conserved pathogenesis-associated proteins: VirB4, VirB9, and VirB10. VirB9 is an OM component of the T4SS and is hypothesized to be a translocation pore (69, 70). It is essential for the stability of the translocation machinery and substrate selection (70). It interacts with VirB10, which bridges the IM and OM protein subcomplexes, and actively participates in T4SS substrate transfer across the bacterial envelope (71-73). VirB4 is an ATPase, providing energy for substrate export and pilus biogenesis, and it interacts with several other VirB proteins, such as VirB10 (51). It is not surprising, then, to identify such proteins in the *E. ruminantium* OM fraction. Moreover, a recent

study showed that some T4SS components could be potential vaccine candidate for pathogenic bacteria (50).

We also identified a porin (ERGA_CDS_04580) that has no homology to other proteins and that seems to be unique to *E. ruminantium*. Porins play a fundamental role in pathogenicity (52), participating in adhesion to and invasion of host cells and evasion of host defense mechanisms (74). They represent good targets for therapeutic development. Some porins activate immunological responses, induce signaling pathways, and modify the properties of the OM lipid barrier (74). It would be interesting to further investigate the role of this porin with functional studies.

The periplasmic major ferric iron binding protein of Gram-negative bacteria (ERGA_CDS_01230), which has homologous counterparts in many other pathogenic species, plays a key role in the acquisition of iron from mammalian host serum iron transport proteins; thus, it is essential for the survival of the pathogen within the host (42, 75).

Within the cell, the full-length protease (ERGA_CDS_06350), may be processed into the intermediate 45 kDa form, which represents a form of protease IV that lacks the signal sequence. This 45 kDa intermediate may undergo a conformational change that activates its protease activity, triggering the cleavage of the propeptide from the mature protease domain. The mature protease IV may be secreted through the OM, functioning in the developmental cycle (76, 77) and as an important virulence factor (78).

In this study, we detected the chaperones DnaK and GroEL in the OM fraction, though they are depicted as cytoplasmic proteins. These results are not surprising,

as these proteins are often membrane-associated (10, 79). In many bacteria, such as *L. pneumophila* and *Borrelia burgdorferi* (9, 80), GroEL (Hsp60) is found in the OM and plays a role in the folding of a large number of proteins; in other bacteria, this protein is active in bacterial adhesion (81, 82). Similarly, in *E. chaffeensis*, the chaperone proteins GroEL and DnaK, and the translation elongation factor G, are localized to the membrane surface (83). GroEL has also been detected on the surfaces of *H. pylori* (84), *L. pneumophila* (85), *Haemophilus ducreyi* (86), and *Clostridium difficile* (82) via immunofluorescence or immunoelectron microscopy. Finally, DnaK has been detected on the surface of *H. pylori* (84). Other important cytoplasmic proteins identified in our study (FusA, TypA, EF-Tu, and Tig) are associated with ribosomes but can be membrane-associated during the transport of nascent OMPs across the periplasmic space to the OM (87). Recently, EF-Tu was shown to be membrane-associated, secreted in outer membrane vesicles (OMVs), and immunogenic during *Burkholderia* infection in a murine model of melioidosis (88). Therefore, we cannot deny the possibility that these proteins with well-known functions in the cytoplasmic, periplasmic, or inner membrane are present in the OM of *E. ruminantium* and play unexpected roles in *E. ruminantium* -host interaction.

Surprisingly, we also detected ribosomal proteins with a predicted cytoplasmic localization. These proteins may represent a contamination with cytoplasmic proteins. Such proteins have also been identified in OM fractions of *Pseudomonas* and *Yersinia* strains, however (89, 90). Moreover, it should be noted that among these ribosomal proteins, we obtained a majority of 50S ribosomal subunits, as has been shown in *Legionella* (9). Interestingly, one ribosomal protein we found in the OM fraction (ERGA_CDS_01640) has been predicted by S4TE software as a

putative type IV effector (27). Type IV effectors are proteins produced by pathogenic bacteria to manipulate host cell gene expression and other processes and have been shown to be critical for pathogenicity, making them salient targets for understanding bacterial virulence (91). The function of this particular protein and its role in *E. ruminantium* pathogenicity is currently under investigation.

Conclusion:

This study provides the first proteomic profile of the *Ehrlichia ruminantium* OM. The combination of subcellular fractionation via sarkosyl solubilization and a high degree of accuracy in predicting OMP status allowed us to generate a high-resolution OM proteome comprised of 46 proteins identified in the OM fraction. We identified OMPs involved in cell wall structure, i.e. at the interface between bacteria and host cells, and proteins known to be virulence factors. Moreover, we identified new OMPs by our approach coupling a consensus of computer algorithms, manual sequence analysis and experimental proteomics. In the future, functional studies should explore the potential of using these OMPs as vaccine candidates against *E. ruminantium*.

Acknowledgments

We are grateful to Elisabete Pires and Renata Soares for technical assistance with MS/MS analysis, and David Pleydell for help with bioinformatics analysis. We thank C3MAG platform for assistance with Transmission Electron Microscopy. The authors acknowledge financial support from FEDER grant FED 1/1.4-30305, 2007–2013, “Risque en santé animale et végétale” (PhD grant to AM), a PAUILF grant (Programme des Actions Universitaires Intégrées Luso-Françaises), 2011–2012, “Identification des protéines de la membrane externe de la *Rickettsiales Ehrlichia ruminantium* par une approche de protéomique,” and the Fundação para a Ciência e

1 Tecnologia (FCT, Lisbon, Portugal; research project PTDC/CVT/114118/2009,
2 research grant PEst-OE/EQB/LA0004/2011, and Post-doc grant SFRH/ BPD/ 45978/
3 2008 to IM).

References

1. **Allsopp BA**. 2010. Natural history of *Ehrlichia ruminantium*. Vet Parasitol **167**:123–135.
2. **Jongejan F**, **Zandbergen TA**, **van de Wiel PA**, **de Groot M**, **Uilenberg G**. 1991. The tick-borne rickettsia *Cowdria ruminantium* has a *Chlamydia*-like developmental cycle. Onderstepoort J. Vet. Res. **58**:227–237.
3. **Vachiery N**, **Marcelino I**, **Martinez D**, **Lefrancois T**. 2013. Opportunities in diagnostic and vaccine approaches to mitigate potential heartwater spreading and impact on the american mainland. Dev Biol (Basel) **135**:191–200.
4. **Jongejan F**, **Vogel SW**, **Gueye A**, **Uilenberg G**. 1993. Vaccination against heartwater using in vitro attenuated *Cowdria ruminantium* organisms. Rev Elev Med Vet Pays Trop **46**:223–227.
5. **Mahan S**, **Smith GE**, **Kumbula D**, **Burridge MJ**, **Barbet A**. 2001. Reduction in mortality from heartwater in cattle, sheep and goats exposed to field challenge using an inactivated vaccine. Vet Parasitol 1–14.
6. **Pretorius A**, **Collins NE**, **Steyn HC**, **van Strijp F**, **van Kleef M**, **Allsopp BA**. 2007. Protection against heartwater by DNA immunisation with four *Ehrlichia ruminantium* open reading frames. Vaccine **25**:2316–2324.
7. **Faburay B**, **Geysen D**, **Ceesay A**, **Marcelino I**, **Alves PM**, **Taoufik A**, **Postigo M**, **Bell-Sakyi L**, **Jongejan F**. 2007. Immunisation of sheep against heartwater in The Gambia using inactivated and attenuated *Ehrlichia ruminantium* vaccines. Vaccine **25**:7939–7947.
8. **Raliniaina M**, **Meyer DF**, **Pinarello V**, **Sheikboudou C**, **Emboulé L**, **Kandassamy Y**, **Adakal H**, **Stachurski F**, **Martinez D**, **Lefrancois T**,

Vachiéry N. 2010. Mining the genetic diversity of *Ehrlichia ruminantium* using map genes family. Vet Parasitol **167**:187–195.

9. **Khemiri A**, **Galland A**, **Vaudry D**, **Chan Tchi Song P**, **Vaudry H**, **Jouenne T**, **Cosette P**. 2008. Outer-membrane proteomic maps and surface-exposed proteins of *Legionella pneumophila* using cellular fractionation and fluorescent labelling. Anal Bioanal Chem **390**:1861–1871.

10. **Rhomberg TA**, **Karlberg O**, **Mini T**, **Zimny-Arndt U**, **Wickenberg U**, **Röttgen M**, **Jungblut PR**, **Jenö P**, **Andersson SGE**, **Dehio C**. 2004. Proteomic analysis of the sarcosine-insoluble outer membrane fraction of the bacterial pathogen *Bartonella henselae*. Proteomics **4**:3021–3033.

11. **Jagannadham MV**, **Abou-Eladab EF**, **Kulkarni HM**. 2011. Identification of Outer Membrane Proteins from an Antarctic Bacterium *Pseudomonas syringae* Lz4W. Molecular & Cellular Proteomics **10**:M110.004549–M110.004549.

12. **Hobb RI**, **Fields JA**, **Burns CM**, **Thompson SA**. 2009. Evaluation of procedures for outer membrane isolation from *Campylobacter jejuni*. Microbiology **155**:979–988.

13. **Ayalew S**, **Confer AW**, **Hartson SD**, **Shrestha B**. 2010. Immunoproteomic analyses of outer membrane proteins of *Mannheimia haemolytica* and identification of potential vaccine candidates. Proteomics **10**:2151–2164.

14. **Shevchuk O**, **Jäger J**, **Steinert M**. 2011. Virulence properties of the *legionella pneumophila* cell envelope. Front. Microbio. **2**:74.

15. **Weiner JH**, **Li L**. 2008. Proteome of the *Escherichia coli* envelope and technological challenges in membrane proteome analysis. BBA Biomembranes **1778**:1698–1713.

16. **Oleastro M**, **Ménard A**. 2013. The Role of *Helicobacter pylori* Outer Membrane Proteins in Adherence and Pathogenesis. Biology **2**:1110–1134.

17. **Marcelino I**, **de Almeida AM**, **Brito C**, **Meyer DF**, **Barreto M**, **Sheikboudou C**, **Franco CF**, **Martinez D**, **Lefrancois T**, **Vachiéry N**, **Carrondo MJT**, **Coelho AV**, **Alves PM**. 2012. Proteomic analyses of *Ehrlichia ruminantium* highlight differential expression of MAP1-family proteins. Vet. Microbiol. **156**:305–314.

18. **Postigo M**, **Taoufik A**, **Bell-Sakyi L**, **Bekker CPJ**, **de Vries E**, **Morrison WI**, **Jongejan F**. 2008. Host cell-specific protein expression in vitro in *Ehrlichia ruminantium*. Vet. Microbiol. **128**:136–147.

19. **Ohashi N**, **Zhi N**, **Zhang Y**, **Rikihisa Y**. 1998. Immunodominant major outer membrane proteins of *Ehrlichia chaffeensis* are encoded by a polymorphic multigene family. Infection and Immunity **66**:132–9.

20. **Lopez JE**, **Palmer GH**, **Brayton KA**, **Dark MJ**, **Leach SE**, **Brown WC**. 2007. Immunogenicity of *Anaplasma marginale* Type IV Secretion System Proteins in a Protective Outer Membrane Vaccine. Infection and Immunity **75**:2333–2342.

21. **Futse, J. E., K. A. Brayton, M. J. Dark, D. P. Knowles, Jr., and G. H. Palmer.** 2008. Superinfection as a driver of genomic diversification in antigenically variant pathogens. Proc. Natl. Acad. Sci. USA **105:**2123–2127.

22. **Felek S**, **Greene R**, **Rikihisa Y**. 2003. Transcriptional analysis of p30 major outer membrane protein genes of *Ehrlichia canis* in naturally infected ticks and sequence analysis of p30-10 of *E canis* from diverse geographic regions. Journal of Clinical Microbiology **41**:886–888.

23. **Unver A**, **Rikihisa Y**, **Stich RW**, **Ohashi N**, **Felek S**. 2002. The omp-1 major

outer membrane multigene family of *Ehrlichia chaffeensis* is differentially expressed in canine and tick hosts. Infection and Immunity **70**:4701–4704.

24. **Filip C**, **Fletcher G**, **Wulff JL**, **Earhart CF**. 1973. Solubilization of the cytoplasmic membrane of *Escherichia coli* by the ionic detergent sodium-lauryl sarcosinate. Journal of Bacteriology **115**:717–722.

25. **Marcelino I**, **Veríssimo C**, **Sousa MFQ**, **Carrondo MJT**, **Alves PM**. 2005. Characterization of *Ehrlichia ruminantium* replication and release kinetics in endothelial cell cultures. Vet. Microbiol. **110**:87–96.

26. **Marcelino I**, **Vachiéry N**, **Amaral AI**, **Roldão A**, **Lefrancois T**, **Carrondo MJT**, **Alves PM**, **Martinez D**. 2007. Effect of the purification process and the storage conditions on the efficacy of an inactivated vaccine against heartwater. Vaccine **25**:4903–4913.

27. **Meyer DF**, **Noroy C**, **Moumène A**, **Raffaele S**, **Albina E**, **Vachiéry N**. 2013. Searching algorithm for type IV secretion system effectors 1.0: a tool for predicting type IV effectors and exploring their genomic context. Nucleic Acids Res **41**:9218–9229.

28. **Ohashi N**, **Unver A**, **Zhi N**, **Rikihisa Y**. 1998. Cloning and characterization of multigenes encoding the immunodominant 30-kilodalton major outer membrane proteins of *Ehrlichia canis* and application of the recombinant protein for serodiagnosis. Journal of Clinical Microbiology **36**:2671–2680.

29. **Gros O**, **Elisabeth NH**, **Gustave SDD**, **Caro A**, **Dubilier N**. 2012. Plasticity of symbiont acquisition throughout the life cycle of the shallow-water tropical lucinid *Codakia orbiculata* (Mollusca: Bivalvia). Environ Microbiol **14**:1584–1595.

30. **Franco CF**, **Santos R**, **Coelho AV**. 2011. Proteome characterization of sea star coelomocytes - The innate immune effector cells of echinoderms. Proteomics **11**:3587–3592.

31. **UniProt Consortium**. 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res **40**:D71–5.

32. **Yu NY**, **Wagner JR**, **Laird MR**, **Melli G**, **Rey S**, **Lo R**, **Dao P**, **Sahinalp SC**, **Ester M**, **Foster LJ**, **Brinkman FSL**. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics **26**:1608–1615.

33. **Yu C-S**, **Lin C-J**, **Hwang J-K**. 2004. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. **13**:1402–1406.

34. **Magnus M**, **Pawlowski M**, **Bujnicki JM**. 2012. MetaLocGramN: A meta-predictor of protein subcellular localization for Gram-negative bacteria. Biochim. Biophys. Acta **1824**:1425–1433.

35. **Matthews BW**. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta **405**:442–451.

36. **Kulkarni HM**, **Swamy CVB**, **Jagannadham MV**. 2014. Molecular characterization and functional analysis of outer membrane vesicles from the antarctic bacterium *Pseudomonas syringae* suggest a possible response to environmental conditions. J. Proteome Res. **13**:1345–1358.

37. **Stekhoven DJ**, **Omasits U**, **Quebatte M**, **Dehio C**, **Ahrens CH**. 2014. Proteome-wide identification of predominant subcellular protein localizations in a bacterial model organism. J Proteomics **99**:123–137.

38. **Roszczenko P**, **Radomska K**, **Wywial E**, **Collet J**, **Jagusztyn-Krynicka E**. 2012. A Novel Insight into the Oxidoreductase Activity of *Helicobacter pylori* HP0231 Protein **7**:e46563.

39. **Lin J**, **Huang S**, **Zhang Q**. 2002. Outer membrane proteins: key players for bacterial adaptation in host niches. Microbes and Infection **4**:325–331.

40. **Wolff S**, **Hahne H**, **Hecker M**, **Becher D**. 2008. Complementary Analysis of the Vegetative Membrane Proteome of the Human Pathogen *Staphylococcus aureus*. Molecular & Cellular Proteomics **7**:1460–1468.

41. **Papanastasiou M**, **Orfanoudaki G**, **Koukaki M**, **Kountourakis N**, **Sardis MF**, **Aivaliotis M**, **Karamanou S**, **Economou A**. 2013. The *Escherichia coli* Peripheral Inner Membrane Proteome. Molecular & Cellular Proteomics **12**:599–610.

42. **Brown RN**, **Romine MF**, **Schepmoes AA**, **Smith RD**, **Lipton MS**. 2010. Mapping the subcellular proteome of *Shewanella oneidensis* MR-1 using sarkosyl-based fractionation and LC-MS/MS protein identification. J. Proteome Res. **9**:4454–4463.

43. **Seo GM**, **Cheng C**, **Tomich J**, **Ganta RR**. 2008. Total, Membrane, and Immunogenic Proteomes of Macrophage- and Tick Cell-Derived *Ehrlichia chaffeensis* Evaluated by Liquid Chromatography-Tandem Mass Spectrometry and MALDI-TOF Methods. Infection and Immunity **76**:4823–4832.

44. **Singu V**, **Liu H**, **Cheng C**, **Ganta RR**. 2005. *Ehrlichia chaffeensis* expresses macrophage- and tick cell-specific 28-kilodalton outer membrane proteins. Infection and Immunity **73**:79–87.

45. **Singu V**, **Peddireddi L**, **Sirigireddy KR**, **Cheng C**, **Munderloh U**, **Ganta RR**.

2006. Unique macrophage and tick cell-specific protein expression from the p28/p30-outer membrane protein multigene locus in *Ehrlichia chaffeensis* and *Ehrlichia canis*. Cell Microbiol **8**:1475–1487.

46. **Koßmehl S**, **Wöhlbrand L**, **Drüppel K**, **Feenders C**, **Blasius B**, **Rabus R**. 2013. Subcellular protein localization (cell envelope) in *Phaeobacter inhibens* DSM 17395. Proteomics **13**:2743–2760.

47. **E-komon T**, **Burchmore R**, **Herzyk P**, **Davies R**. 2012. Predicting the outer membrane proteome of *Pasteurella multocida* based on consensus prediction enhanced by results integration and manual confirmation. BMC Bioinformatics **13**:63.

48. **Peter TF**, **Burridge MJ**, **Mahan S**. 2002. *Ehrlichia ruminantium* infection (heartwater) in wild animals. Trends Parasitol. **18**:214–8.

49. **van den Berg B**. 2013. Lateral gates: β-barrels get in on the act. Nat. Struct. Mol. Biol. **20**:1237–1239.

50. **Morse K**, **Norimine J**, **Palmer GH**, **Sutten EL**, **Baszler TV**, **Brown WC**. 2012. Association and Evidence for Linked Recognition of Type IV Secretion System Proteins VirB9-1, VirB9-2, and VirB10 in *Anaplasma marginale*. Infection and Immunity **80**:215–227.

51. **Ward DV**, **Draper O**, **Zupan JR**, **Zambryski PC**. 2002. Peptide linkage mapping of the *Agrobacterium tumefaciens* vir-encoded type IV secretion system reveals protein subassemblies. Proc. Natl. Acad. Sci. U.S.A. **99**:11493–11500.

52. **Achouak W**, **Heulin T**, **Pagès JM**. 2001. Multiple facets of bacterial porins. FEMS Microbiology Letters **199**:1–7.

53. **Nowalk AJ**, **Burroughs Tencza S**, **Mietzner TA**. 1994. Coordination of Iron by the Ferric Iron-Binding Protein of Pathogenic *Neisseria* Is Homologous to the Transferrins. Biochemistry **33**:12769–12775.

54. **Ge Y**, **Rikihisa Y**. 2007. Surface-Exposed Proteins of *Ehrlichia chaffeensis*. Infection and Immunity **75**:3833–3841.

55. **van Heerden H**, **Collins NE**, **Brayton KA**, **Rademeyer C**, **Allsopp BA**. 2004. Characterization of a major outer membrane protein multigene family in *Ehrlichia ruminantium*. Gene **330**:159–168.

56. **Bekker CPJ**, **Postigo M**, **Taoufik A**, **Bell-Sakyi L**, **Ferraz C**, **Martinez D**, **Jongejan F**. 2005. Transcription analysis of the major antigenic protein 1 multigene family of three in vitro-cultured *Ehrlichia ruminantium* isolates. Journal of Bacteriology **187**:4782–4791.

57. **Frutos R**, **Viari A**, **Ferraz C**, **Morgat A**, **Eychenié S**, **Kandassamy Y**, **Chantal I**, **Bensaid A**, **Coissac E**, **Vachiery N**. 2006. Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. Journal of Bacteriology **188**:2533–2542.

58. **Tellez R**, **Misra R**. 2012. Substitutions in the BamA β-barrel domain overcome the conditional lethal phenotype of a Δ*bamB* Δ*bamE* strain of *Escherichia coli*. Journal of Bacteriology **194**:317–324.

59. **Nikaido H**. 2003. Molecular Basis of Bacterial Outer Membrane Permeability Revisited N2 . Microbiology and Molecular Biology Reviews 1–64.

60. **Rigel NW**, **Ricci DP**, **Silhavy TJ**. 2013. Conformation-specific labeling of BamA and suppressor analysis suggest a cyclic mechanism for β-barrel assembly in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **110**:5151–5156.

61. **Werner J**, **Misra R**. 2005. YaeT (Omp85) affects the assembly of lipid-dependent and lipid-independent outer membrane proteins of *Escherichia coli*. Mol Microbiol **57**:1450–1459.

62. **Doerrler WT**, **Raetz CRH**. 2005. Loss of Outer Membrane Proteins without Inhibition of Lipid Export in an *Escherichia coli* YaeT Mutant. J. Biol. Chem. **280**:27679–27687.

63. **Hoang HH**, **Nickerson NN**, **Lee VT**, **Kazimirova A**, **Chami M**, **Pugsley AP**, **Lory S**. 2011. Outer Membrane Targeting of *Pseudomonas aeruginosa* Proteins Shows Variable Dependence on the Components of Bam and Lol Machineries. MBio **2**:e00246–11–e00246–11.

64. **Malinverni JC**, **Werner J**, **Kim S**, **Sklar JG**, **Kahne D**, **Misra R**, **Silhavy TJ**. 2006. YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. Mol Microbiol **61**:151–164.

65. **Sklar JG**, **Wu T**, **Gronenberg LS**, **Malinverni JC**, **Kahne D**, **Silhavy TJ**. 2007. Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **104**:6400–6405.

66. **Voth D**, **Broederdorf L**, **Graham J**. 2013. Bacterial Type IV secretion systems: versatile virulence machines. Future Microbiol. **7**:241-257.

67. **Christie PJ**. 2009. Structural biology: Translocation chamber's secrets. Nature **462**:992–994.

68. **Dehio C**. 2008. Infection-associated type IV secretion systems of *Bartonella* and their diverse roles in host cell interaction. Cell Microbiol **10**:1591–1598.

69. **Cascales E**, **Christie PJ**. 2003. The versatile bacterial type IV secretion

systems. Nat Rev Micro **1**:137–149.

70. **Jakubowski SJ**, **Cascales E**, **Krishnamoorthy V**, **Christie PJ**. 2005. *Agrobacterium tumefaciens* VirB9, an Outer-Membrane-Associated Component of a Type IV Secretion System, Regulates Substrate Selection and T-Pilus Biogenesis. Journal of Bacteriology **187**:3486–3495.

71. **Fronzes R**, **Christie PJ**, **Waksman G**. 2009. The structural biology of type IV secretion systems. Nat Rev Micro **7**:703–714.

72. **Alvarez-Martinez CE**, **Christie PJ**. 2009. Biological diversity of prokaryotic type IV secretion systems. Microbiol. Mol. Biol. Rev. **73**:775–808.

73. **Cascales E**, **Christie PJ**. 2004. Definition of a Bacterial Type IV Secretion Pathway for a DNA Substrate. Science **304**:1170–1173.

74. **Galdiero S**, **Falanga A**, **Cantisani M**, **Tarallo R**, **Pepa Della ME**, **D'Oriano V**, **Galdiero M**. 2012. Microbe-host interactions: structure and role of Gram-negative bacterial porins. Curr. Protein Pept. Sci. **13**:843–854.

75. **Shouldice SR**, **Dougan DR**, **Skene RJ**, **Tari LW**, **McRee DE**, **Yu R-H**, **Schryvers AB**. 2003. High resolution structure of an alternate form of the ferric ion binding protein from *Haemophilus influenzae*. J. Biol. Chem. **278**:11513–11519.

76. **Traidej M**. 2003. Molecular Analysis of *Pseudomonas aeruginosa* Protease IV Expressed in Pseudomonas putida. Investigative Ophthalmology & Visual Science **44**:190–196.

77. **Hoge R**, **Pelzer A**, **Rosenau F**, **Wilhelm S**. 2010. Weapons of a pathogen: proteases and their role in virulence of *Pseudomonas aeruginosa*, pp. 383–395. *In* In Current Research, Technology and Education Topics in Applied

Microbiology and Microbial BiotechnologyMicrobiology book series.

78. **Engel LS**. 1998. Protease IV, a Unique Extracellular Protease and Virulence Factor from *Pseudomonas aeruginosa*. Journal of Biological Chemistry **273**:16792–16797.

79. **Boonjakuakul JK**, **Gerns HL**, **Chen YT**, **Hicks LD**, **Minnick MF**, **Dixon SE**, **Hall SC**, **Koehler JE**. 2007. Proteomic and Immunoblot Analyses of *Bartonella quintana* Total Membrane Proteins Identify Antigens Recognized by Sera from Infected Patients. Infection and Immunity **75**:2548–2561.

80. **Scopio A**, **Johnson P**, **Laquerre A**, **Nelson DR**. 1994. Subcellular localization and chaperone activities of *Borrelia burgdorferi* Hsp60 and Hsp70. Journal of Bacteriology **176**:6449–6456.

81. **Kusukawa N**, **Yura T**, **Ueguchi C**, **Akiyama Y**, **Ito K**. 1989. Effects of mutations in heat-shock genes *groES* and *groEL* on protein export in *Escherichia coli*. The EMBO Journal **8**:3517–3521.

82. **Hennequin C**, **Collignon A**, **Karjalainen T**. 2001. Analysis of expression of GroEL (Hsp60) of *Clostridium* difficile in response to stress. Microbial Pathogenesis **31**:255–260.

83. **Ge Y**, **Rikihisa Y**. 2007. Identification of novel surface proteins of *Anaplasma phagocytophilum* by affinity purification and proteomics. Journal of Bacteriology **189**:7819–7828.

84. **Huesca M**, **Borgia S**, **Hoffman P**, **Lingwood CA**. 1996. Acidic pH changes receptor binding specificity of *Helicobacter pylori*: a binary adhesion model in which surface heat shock (stress) proteins mediate sulfatide recognition in gastric colonization. Infection and Immunity **64**:2643–2648.

85. **Garduño RA**, **Garduno E**, **Hoffman PS**. 1998. Surface-associated *hsp60* chaperonin of *Legionella pneumophila* mediates invasion in a HeLa cell model. Infection and Immunity **66**:4602–4610.

86. **Frisk A**, **Ison CA**, **Lagergard T**. 1998. GroEL heat shock protein of *Haemophilus ducreyi*: association with cell surface and capacity to bind to eukaryotic cells. Infection and Immunity **66**:1252–1257.

87. **Herskovits AA**. 2002. Accumulation of endoplasmic membranes and novel membrane-bound ribosome-signal recognition particle receptor complexes in *Escherichia coli*. J. Cell Biol. **159**:403–410.

88. **Nieves W, Heang J, Asakrah S, Höner zu Bentrup K, Roy CJ, Morici LA.** 2010. Immunospecific responses to bacterial elongation factor Tu during *Burkholderia* infection and immunization. PLoS ONE **5**:e14361.

89. **Seyer D**, **Cosette P**, **Siroy A**, **Dé E**, **Lenz C**, **Vaudry H**, **Coquet L**, **Jouenne T**. 2005. Proteomic comparison of outer membrane protein patterns of sessile and planktonic *Pseudomonas aeruginosa* cells. Biofilms **2**:27–36.

90. **Coquet L**, **Cosette P**, **Dé E**, **Galas L**, **Vaudry H**, **Rihouey C**, **Lerouge P**, **Junter G-A**, **Jouenne T**. 2005. Immobilization induces alterations in the outer membrane protein pattern of *Yersinia ruckeri*. J. Proteome Res. **4**:1988–1998.

91. **Hicks SW**, **Galan JE**. 2013. Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. Nat Rev Micro **11**:316–326.

Figure Legends

Figure 1. Experimental workflow for *E. ruminantium* subcellular fractionation and proteome characterization. **OM**, outer membrane; **I**, inner membrane; **C**, cytoplasm.

Figure 2. Evaluation of OM isolation quality. Transmission electron microscopy of (A) purified *E. ruminantium* and (B) the insoluble precipitate after 0.1% sarkosyl treatment; scale bar = 200 nm. (C) SDS-PAGE and (D) Western blot of **E** (elementary bodies), **S** (sarkosyl-soluble fraction), and **OM** (outer membrane fraction) using monoclonal antibodies against Map1. **Band 1:** Map1-14, X5HG56, GroEL; **Band 2:** Map1+1, Map1, Map1-6, VirB10, VirB4, GroEL, PyrE, Q5HAR6, X5HG56, 30S-S8; **Band 3:** Map1+1, Map1-6, Map2, GroEL, PyrE, Q5HAR6, X5HG56, Q5FGC2, Q5HBI2, Q5FHJ9; **Band 4:** Map1, Map1-6, VirB4, GroEL, DnaK, BamA, FusA, Pnp, Q5HAR6, X5HG56, Q5FH07, Q5HBS6; **Band 5:** VirB4, VirB10, VirB11, DnaK, HtpG, GroEL, FusA, 30S-S1, Q5FGV5, Q93FS2; **Band 6:** Map1, Map1-14, VirB10, PleD, GroEL, DnaK, FtsZ, 30S-S1, Q5HB83, Q5FGA7, Q5HBE1; **Band 7:** Map1-14, GroEL, DnaK, FtsZ, HtpG; **Band 8:** Map1-6, Map1, GroEL, DnaK, FtsZ, BamA; **Band 9:** Map1-6, Map1, Map1+1, Map1-14,GroEL, DnaK, BamA, Q5FFE6, Q5HAR6; **Band 10:** Map1-11, Map1-13, Map1, Map1+1, Map1-6, VirB10, VirB9, Q5FFE6, Q5HAR6, Q5HBI2, Q5HA95; **Band 11:** Map1, Map2, BamA, DnaK, GroEL, FusA, Def, 50S-L4, PyrE, X5HG56, Q5HBI2; **Band 12:** 30S-S18, 30S-S12, 50S-L7/L12, 50S-L18, 50S-L24, 50S-L28 X5HG56, Q5HBN6; **Band 13:** HupB, X5HG56; **Band 14:** 30S-S12, 50S-L7/L12, 50S-L18, GroEL, YajC, PyrE

Figure 3. Venn diagram representing the predicted subcellular localization of *E. ruminantium* proteins using PSORTb 3.0, CELLO 2.5, and MetaLocGramN. The data presented result from consensus prediction of subcellular localization.

Table 1. Subcellular localization of *E. ruminantium* strain Gardel proteins as predicted by PSORTb 3.0, CELLO 2.5, MetalocGramN, and consensus. Percentages correspond to the number of proteins in each compartment relative to the total number of proteins.

Subcellular localization	PSORTb 3.0		CELLO 2.5		MetaLocGramN		Consensus prediction	
	Number	%	Number	%	Number	%	Number	%
Cytoplasmic	490	51.6	461	48.6	526	55.4	499	52.6
Periplasmic	4	0.4	9	0.9	1	0.1	1	0.1
Inner Membrane	198	20.8	109	11.4	192	20.2	124	13.0
Extracellular	9	0.9	23	2.4	158	16.6	16	1.6
Outer membrane	11	1.1	90	9.4	71	7.4	52	5.4
Unknown	236	24.8	256	27.0	0	0	256	27.0
Total	948		948		948		948	

Table 2. Proteins identified in the outer membrane fraction of *E. ruminantium* via 1DE-nanoLC-MALDI-TOF/TOF. Their predicted subcellular localization is shown by U, unknown; C, cytoplasmic; I, inner membrane; O, outer membrane; E, extracellular; P, periplasmic.

Locus tag	Protein	Acession Number	Function	Protein MM (kDa)	Number peptides[a]	Protein score[b]	Coverage (%)	PSORTb 3.0	CELLO 2.5		MetaLocGramN	Consensus prediction
Outer membrane proteins (39%)												
ERGA_CDS_00150	VirB10	Q5HCE9	Virulence	48.717	2	294	9	I	E	P C	C	OM
ERGA_CDS_00160	VirB9-2	Q5HCE8	Virulence	30.993	2	86	6	C	OM		C	OM
ERGA_CDS_01230	Possible major ferric iron binding protein	Q5FFA9	Transport/virulence	41.309	2	188	14	OM	C	OM	I	OM
ERGA_CDS_02370	Hypothetical protein	Q5FFH4	Unknown	37.402	3	259	26	U	OM		E	OM
ERGA_CDS_02510*	Hypothetical protein	Q5HBS6	Unknown	90.496	1	26	1	U	OM		OM	OM
ERGA_CDS_03960*	Hypothetical protein	Q5HBE1	Unknown	55.237	1	23	2	U	OM		E	OM
ERGA_CDS_04510	Hypothetical protein	Q5FGV5	Unknown	134.574	1	124	2	U	C	OM	OM	OM
ERGA_CDS_04580	Putative exported protein	Q5HB83	Porin	41.826	9	832	47	U	OM		OM	OM
ERGA_CDS_05150[#]	Putative exported protein	Q5FH07	Unknown	63.139	18	827	27	OM	OM		OM	OM
ERGA_CDS_07300	Hypothetical outer membrane protein	Q93FS2	Cell struture	28.127	2	107	18	U	C		E	OM
ERGA_CDS_07840	VirB9-1	Q5HAC9	Virulence	29.489	2	153	10	I	C	OM	C	OM
ERGA_CDS_08100	Putative exported lipoprotein	Q5HAA5	Outer membrane assembly	29.344	1	119	10	OM	C	OM	OM	OM
ERGA_CDS_08660	Outer membrane protein omp1	Q5FGI9	Outer membrane assembly	87.257	2	173	5	OM	OM		OM	OM
ERGA_CDS_09000	Map1-13	Q4L0D3	Cell struture	32.965	4	419	28	I	OM	P	C	OM
ERGA_CDS_09010	Map1-14	Q4W4X7	Cell struture	34.186	1	122	7	U	OM		OM	OM
ERGA_CDS_09090	Map1-6	Q4L0C5	Cell struture	33.736	4	529	36	OM	OM		OM	OM
ERGA_CDS_09160	Map1	Q46330	Cell structure	31.204	5	948	38	OM	E		E	OM
ERGA_CDS_09170	Map1+1	Q4L0B8	Cell structure	31.817	3	136	14	U	OM		OM	OM
Inner membrane proteins (11%)												
ERGA_CDS_01470	Major antigenic protein 2 SCO2 like-protein	Q9R416	Cell struture	23.562	1	79	10	I	P	C	I	I
ERGA_CDS_03170	Phosphatidylserine decarboxylase proenzyme	Q5FHJ9	General metabolism	25.245	1	103	7	I	I		I	I
ERGA_CDS_05400	VirB4	Q5FFK8	Virulence	90.865	1	82	2	I	C		C	I
ERGA_CDS_06350	Putative Protease IV	Q5HAR6	General metabolism	32.263	2	157	14	I	C	OM	I	I
ERGA_CDS_08130	Preprotein translocase. YajC subunit	X5HHA7	Cellular processes and signaling	13	1	43	15	I	C	P	I	I
Cytoplasmic proteins (50%)												
ERGA_CDS_01570	Elongation factor G	Q5FFE7	Protein synthesis	76.042	1	40	3	C	C		C	C
ERGA_CDS_01580	Elongation factor Tu	Q5FFE6	Protein synthesis	43.282	5	344	17	C	C		C	C
ERGA_CDS_01760	Peptide deformylase	Q5HBZ5	Cell process	21.926	1	43	9	C	C		C	C
ERGA_CDS_02930	Putative DNA-binding protein HU-beta	Q5HBN6	Cell process	10.665	1	94	19	U	C	P	E	C
ERGA_CDS_03000[#]	Helix-turn-helix domain protein	X5HG56	DNA binding	12 250	1	40	6	U	C		E	C
ERGA_CDS_03230	Response regulator pleD	Q5HBK9	Regulation	52.358	1	154	5	C	C		C	C
ERGA_CDS_03510	Putative peroxiredoxin	Q5HBI2	General metabolism	23.349	8	760	52	C	C		C	C
ERGA_CDS_03570	Polyribonucleotide nucleotidyltransferase	Q5FHK5	General metabolism	86.507	1	77	3	C	OM		C	C
ERGA_CDS_07810	Inosine-5'-monophosphate dehydrogenase	Q5FGA7	General metabolism	52.348	1	106	6	C	C		C	C
ERGA_CDS_08210	Putative response regulator	Q5HA95	Regulation	30.477	2	118	9	C	C		C	C
ERGA_CDS_08900	Orotate phosphoribosyltransferase	Q5FGJ6	General metabolism	22.343	1	57	7	C	C		C	C

Partie 2

Etude de la régulation du SST4 et des protéines Map1 de la membrane externe d'*E. ruminantium*

1. Préambule

Pour infecter son hôte animal, *E. ruminantium* doit réguler finement l'ensemble des fonctions liées à la pathogénèse de manière à adapter spatialement et temporellement l'expression génique en fonction des conditions environnementales. Les paralogues de la famille multigénique *map1* d'*E. ruminantium* codent pour 16 protéines de la membrane externe dont Map1 qui est une protéine antigénique majeure. Cette famille est homologue à des familles multigéniques similaires chez *Ehrlichia chaffeensis* (*p28*) et chez *Anaplasma* (*msp2, p44*). Bien que leur rôle précis soit encore inconnu, les protéines de la famille Map1 semblent essentielles à la croissance intracellulaire d'*E. ruminantium* (Pruneau et al. 2014).

Les bactéries intracellulaires utilisent les systèmes de sécrétion pour infecter, proliférer et persister à l'intérieur de leur hôte (Tseng, Tyler, and Setubal 2009). Les signaux et les mécanismes de régulation de ces systèmes de sécrétion sont largement étudiés (Baumgartner et al. 2014; Miyata, Bachmann, and Pukatzki 2013; Wu et al. 2012). Des études antérieures ont montré que le SST4 chez *Ehrlichia chaffeensis* n'était pas exprimé de façon constitutive mais finement régulé par des facteurs de transcription (Z. Cheng, Wang, and Rikihisa 2008). Ainsi, le facteur de transcription EcxR se lie aux promoteurs des gènes *vir* codant pour le système de sécrétion de type IV (SST4) et induit leur expression aux stades précoces du cycle de développement (Z. Cheng, Wang, and Rikihisa 2008). De plus, chez *A. phagocytophilum*, ApxR, l'homologue d'EcxR, induit l'expression des gènes de la famille multigénique *p44* (homologue à la famille *map1*) et du facteur de transcription *tr1* en amont du cluster *p44* (Xueqi Wang et al. 2007; Xueqi Wang, Kikuchi, and Rikihisa 2007).

Dans cette étude, nous avons identifié un orthologue d'EcxR que nous avons nommé ErxR par analogie pour « *E. ruminantium* expression regulator ». Nous avons étudié comment cette protéine était impliquée dans la régulation des gènes *vir et* des gènes de la famille *map1* via facteur de transcription *tr1* chez *E. ruminantium*. Nos résultats montrent qu'ErxR se lie et régule certains promoteurs des gènes *vir*, le promoteur *tr1* en amont des gènes *map1* et lui-même. En se liant à *tr1*, il est donc possible qu'ErxR régule l'expression des gènes de la famille *map1* de façon similaire à celle dont *tr1* régule l'expression du locus *p44*. Dans un

second temps, nous avons voulu déterminer quels étaient les signaux environnementaux influant sur l'expression des gènes *vir* et des gènes de la famille *map1*.

Les résultats obtenus montrent une co-induction *in vitro* des gènes *vir*, *map*, et *erxR* par un pH acide et une carence en Fer, mimant ainsi les conditions rencontrées dans la vacuole. De plus, nos résultats semblent être en faveur d'une co-régulation des gènes *vir, tr1* et *map* par ErxR, ce qui confirme les résultats obtenus chez *E. chaffeensis* pour la régulation du SST4 et suggère pour une première fois un lien direct entre les protéines Map et la virulence.

Ces résultats permettent de mieux comprendre comment les bactéries du genre *Ehrlichia* régulent certains déterminants majeurs du pouvoir pathogène en réponse à des signaux environnementaux. De plus, cette étude permet de commencer à discerner comment la bactérie perçoit sa cellule hôte et déclenche le processus infectieux. L'exploitation de ces résultats pourrait permettre *in fine* la génération de nouveaux moyens de lutte efficaces contre la maladie.

2. Manuscrit préliminaire: Acidic and iron starvation conditions upregulate *Ehrlichia ruminantium* Type IV secretion system and *map1* genes through the master regulatory protein ErxR

Acidic and iron starvation conditions upregulate *Ehrlichia ruminantium* Type IV secretion system and *map1* genes through the master regulatory protein ErxR

Amal Moumène[a,b,c,1], Adela Oliva Chavez[a,b,1], Silvina Gonzalez-Rizzo[d], Thierry Lefrançois[b,e], Nathalie Vachiéry[a,b] and Damien F. Meyer[a,b,*]

[a] CIRAD, UMR CMAEE, Site de Duclos, Prise d'eau, F-97170 Petit-Bourg, Guadeloupe, France
[b] INRA, UMR1309 CMAEE, F-34398 Montpellier, France
[c] Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France
[d] UMR 7138, équipe Biologie de la Mangrove, Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France
[e] CIRAD, UMR CMAEE, Montpellier, France

[*]Corresponding author: damien.meyer@cirad.fr

[1]These authors contributed equally to this work.

Acidic and iron starvation conditions upregulate *Ehrlichia ruminantium* Type IV secretion system and *map1* genes through the master regulatory protein ErxR

Amal Moumène[a,b,c,1], Adela Oliva Chavez[a,b,1], Silvina Gonzalez-Rizzo[d], Thierry Lefrançois[b,e], Nathalie Vachiéry[a,b] and Damien F. Meyer[a,b,*]

[a] CIRAD, UMR CMAEE, Site de Duclos, Prise d'eau, F-97170 Petit-Bourg, Guadeloupe, France

[b] INRA, UMR1309 CMAEE, F-34398 Montpellier, France

[c] Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

[d] UMR 7138, équipe Biologie de la Mangrove, Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

[e] CIRAD, UMR CMAEE, Montpellier, France

[*]Corresponding author: damien.meyer@cirad.fr

[1]These authors contributed equally to this work.

Abstract

Ehrlichia ruminantium is an obligatory intracellular bacterium, transmitted by ticks of the genus *Amblyomma* that causes heartwater, a fatal disease of ruminants. *E. ruminantium* belongs to the *Rickettsiales* and uses a type IV secretion system to deliver effector proteins inside mammalian host cell and subvert host immune responses. Moreover, the genome of *E. ruminantium* contains 16 paralogs of *map1* genes family that are the primary bacterial surface-exposed antigens recognized by the host immune system. Several studies were conducted to understand T4SS regulation or try to decipher the function of Map proteins in *Ehrlichia*, but no integrated approach of the regulation of *Ehrlichia* pathogenicity determinants in response to the environment was carried out. Vacuole or host cell acidification serves as cues for the induction of virulence factors such as type IV or type III secretion systems in several pathogenic bacteria. Iron is a key nutrient for bacterial growth in the environment and within hosts. In this study, we performed transcriptional analyses to compare the expression of *E. ruminantium* T4SS and *map1* genes in standard versus iron depleted or acidic conditions. We showed that exposure of *E. ruminantium* to acidic pH or iron starvation induces ErxR-dependent expression of T4SS apparatus and *map1* genes. We revealed for the first time that Map1 proteins are linked to virulence and might be involved, directly or indirectly, in the perception of environmental signals. Moreover, this study suggests that ErxR, as a master regulatory protein that enables the integration of environmental and nutrient cues, is a key pathogenicity determinant. Altogether, our results show a tight co-regulation of T4SS and *map1* genes at the transcriptional level and thus contribute to a better understanding of the infection process of *Ehrlichia*. Deciphering how *Ehrlichia* senses its environment and subsequently regulates its pathogenicity will provide valuable insights towards the development of alternative therapeutic strategies against these bacteria.

Introduction

Pathogenic bacteria have evolved numerous strategies to manipulate and evade the host's immune responses and to hijack host cells cellular processes in order to survive and proliferate. They developed specialized multiprotein complexes, namely secretion systems, that transport proteins or DNA across their membranes to the extracellular milieu or the host cells in response to specific environmental cues (Tseng et al., 2009). One of such secretion systems is the type IV secretion system (T4SS). T4SS has been shown to be critical for pathogenicity of many bacteria, such as *Bordetella pertussis*, *Helicobacter pylori*, *Legionella pneumophila*, *Coxiella burnetii*, *Bartonella* spp., *Brucella* spp., and *Rickettsial* spp. (Cascales and Christie, 2003; Llosa et al., 2009; Terradot and Waksman, 2011) making it a potential drug targets for the development of new anti-bacterial molecules (Baron and Coombes, 2007). The T4SS is well conserved in the *Anaplasmataceae* family and several T4SS effectors (T4Es) have been described and shown to play an important role in invasion and pathogenesis in members of this family (IJdo et al., 2007; Niu et al., 2010; Lockwood et al., 2011; Liu et al., 2012). *Ehrlichia ruminantium*, the causative agent of heartwater, a fatal disease of ruminants in sub-Saharan African and tropics, belongs to the *Anaplasmataceae* family and is transmitted by ticks of the genus *Amblyomma* (Dumler et al., 2001; Allsopp, 2010). In the mammalian host, *E. ruminantium* replicates primarly within endothelial cells forming membrane-bound vacuole where they replicate (Zweygarth and Josemans, 2001). The biogenesis of this replicative niche depends on the function of T4SS but little is known about T4SS in *E. ruminantium* and no T4Es have been characterized yet (Collins et al., 2005). In several bacteria, it has been shown that the expression of T4SS is not constitutive but tightly regulated by transcription factors (Li and Carlow, 2012; Martín-Martín et al., 2012). Cheng *et al.* (Z. Cheng et al., 2008) showed that the five *virB/D4* genetic loci of *E. chaffeensis* T4SS are co-regulated by the transcription factor EcxR to allow developmental stage-specific expression. In addition, genes encoding the T4SS of *A. phagocytophilum* are also co-regulated during *A. phagocytophilum* growth in human peripheral blood neutrophils (Niu et al., 2008). A EcxR homolog, ApxR, has been found in several *A. phagocytophilum* strains (Z. Cheng et al., 2008). ApxR regulates the expression of the transcription factor *tr1* (X. Wang, Kikuchi, et al., 2007) and the

downstream *p44E* locus (X. Wang, Cheng, et al., 2007). The transcriptional regulatory gene *tr1* has the strongest activity on the *p44/msp2* locus in *A. phagocytophilum* and *A. marginale* (Barbet, Gene 2005). *p44E* encode a large family of major surface proteins whose expression varies depending of the host cell type in which *A. phagocytophilum* is developing, tick or mammalian host cells (Jauron et al., 2001). A homolog of EcxR and ApxR was found in *E. ruminantium* and *tr1* is present upstream of the outer membrane proteins (OMPs) locus *map1,* which are homolog to the *p44* gene family (Rikihisa, 2010). In *E. ruminantium*, the sixteen paralogs of the *map1* multigene family are expressed in bovine endothelial cells and some are preferentially transcribed in the tick or in the mammalian host (van Heerden et al., 2004). Interestingly, the *virB-virD4* loci of *A. phagocytophilum*, *E. chaffensis* and *E. canis* are flanked by genes encoding OMPs members of the P44 family, which are paralogs of MAP1 (Dunning Hotopp et al., 2006). The localization of the genes may reflect a coordination of expression and function among T4SS and various outer membrane proteins.

Microorganisms have evolved elaborate sensory mechanisms in order to regulate their cellular activities in response to environmental changes. This is particularly true for bacterial pathogens whose expression of virulence factors is tightly regulated in response to host and non-host environments (Hyytiäinen et al., 2003). Thus, the regulation of the T4SS in response to host cues can allow for efficient utilization of bacterial resources and facilitate colonization, leading to full infection (Abromaitis et al., 2013). One environmental signal is iron, which is an essential nutrient used in various enzymatic reactions like respiration, DNA replication, oxygen transport, oxidative stress response but can be toxic at elevated intracellular concentrations (Andrews et al., 2003). Therefore, iron scavenging from the limited free iron sources available in the host is a critical bacterial pathogenicity determinant (Ratledge and Dover, 2000). Pathogens have evolved ways to scavenge iron from the host, including the expression of iron acquisition genes as response to low iron concentrations (Brickman et al., 2011; Portier et al., 2014). For example, Fur is a transcriptional regulator that represses the expression of iron acquisition genes when iron is at elevated intracellular concentration (Bagg and Neilands, 1987; Lee and Helmann, 2007). Under conditions of iron starvation, the Fur repression is abolished

and genes involved in iron uptake are transcribed (de Lorenzo et al., 1987). Microbial iron starvation is thus an important signal that controls the expression of other known virulence factors like secretion systems (Mekalanos, 1992). In Gram-negative bacteria, TonB-dependent outer membrane receptors (TBDR) are required for transfer of iron chelates and heme to the periplasm, followed by transport to the cytoplasm (Shultis et al., 2006). The *map1* genes could play a similar role in *E. ruminantium*. Members of the *Anaplasmataceae* family enter into the host cell by the process of endocytosis. After entry, a vacuole, or morula, is formed by the interaction of bacterial proteins with host cell proteins. This vacuole resembles an early autophagosome, presenting a double lipid layer and several protein markers that are characteristic of autophagosomes, including Beclin-1 (Niu and Rikihisa, 2013) Arrest of the vacuole in the early autophagosome inhibits lysosomal fusion and autophagy, leading to the survival of the bacteria (Niu and Rikihisa, 2013). Autophagosomes are acid compartments and mammalian cells require this internal acidification for vesicular transport (Clague et al., 1994). Vesicular transport and induction of autophagosomes are necessary for *Anaplasma phagocytophilum* growth (Niu and Rikihisa, 2013). Thus, it is possible that acidification of the morulae is a conserved mechanism in the *Anaplasmataceae,* which may serve as a clue for the induction of virulence factors, such as the T4SS and T4SS effectors, as seen in other bacteria (Rappl et al., 2003). Indeed, acidification of the infected host cell triggers the expression of the type III secretion system in *Salmonella typhimurium* (Arpaia et al., 2011).

This study was conducted to evaluate the role of pH and iron starvation in the expression of the *E. ruminantium* T4SS and *map1* genes and to determine if the protein ErxR is involved in regulation of these gene clusters. The results indicate that the regulator ErxR plays major role in the expression of T4SS and *map1* genes in response to environmental signals and suggest for the first time that *map1* protein may be important in sensing the environmental signals and are linked to virulence *sensu stricto*. Understanding *Ehrlichia* genes regulation in response to environmental signals during host cell entry can provide valuable clues towards production of an effective vaccine and reveal much about the infectious process of this bacterium.

Materials and methods

Culture conditions

E. ruminantium Gardel strain was routinely propagated in bovine aortic endothelial (BAE) cells as previously described (Marcelino et al., 2005). To evaluate the growth characteristics of *E. ruminantium* at acidic pH or under iron depletion conditions, the strain was grown in BHK-21 cell medium supplemented with 2 mM glutamine, 10% heat inactivated fetal bovine serum (FBS), penicillin (100 IU/ml), streptomycin (100 mg/ml). For expression experiments, the pH of the medium was adjusted to 6 or 7 with 2,5 N of hydrochloric acid, or supplemented with iron (100 µM $FeSO_4$) or an iron chelator (100 µM 2,2'-bipyridyl; BPD) as described in (Breuer et al., 1995; Romeo et al., 2001). The cells were kept in a humidified atmosphere supplemented with 5% CO2 at 37°C. Hydrochloridic acid, $FeSO_4$ or 2,2'-bipyridyl was added when 80% cell lysis was observed, at 120 h post-inoculation (hpi). Twenty four hours after chemical was added, the cell monolayer was harvested by trypsinization and 1/10 (600 µl) of sample was collected by centrifugation at 14,000 X *g* for 10 min. The pellet was kept at -80°C for DNA extraction. The remaining 9/10 (5400 µl) infected cells were centrifuged at 20,000 X *g* for 10 min. The pellet was resuspended in TRIzol reagent (Invitrogen) and stored at -80°C for RNA extraction (Pruneau et al., 2012).

Quantitative detection of *E. ruminantium*

Genomic DNA was extracted from the 1/10 samples, using QIAamp DNA Mini Kit (Qiagen, France). The number of *Ehrlichia ruminantium* per sample was quantified by q-PCR, targeting the single copy of *map1* gene and encodes for a major antigenic protein. The primer sequences are shown in Table 1. A standard curve was established using gDNA of Gardel serially diluted from 7 X 10^6 to 7 X 10^1 copies $\mu L^{-}1$, to determine the number of bacteria per microliter (Pruneau et al., 2012). Four microliters were added to Taqman master mix (Applied Biosystems, France), following the manufacturer's instructions. The PCR conditions were as follow: 2 min at 50 °C, 10 min at 95 °C, and 40 cycles with 15 s at 95 °C and 1 min at 60 °C.

Relative gene expression: RNA preparation and RT-PCR

Total RNA was extracted using TRIzol reagent. RNA pellets were dissolved in 100 µl of DEPC water and treated with turboDNAse (Ambion, France). Isolated RNA purity

and concentration were assessed using a Nanodrop 2000c (Thermo Scientific, France). RNA samples were diluted in RNase-free water at a final concentration of 0.5 µg/µL. RNA samples were reverse-transcribed using SuperScript VILO cDNA Synthesis kit (Invitrogen, France), according to the manufacturer's instructions. The *recA* gene was used as normalizer (Table 1). Quantitative PCR was performed in a 7500 Real-Time PCR System (Applied Biosystems, France) using Power SYBR Green PCR Master Mix (Applied Biosystems, France). Reactions were performed in 25 µl volume with 5 ng template cDNA and 5 µM of each primer. The conditions for the amplification were as follow: 2 min at 50°C, 10 min at 95°C and 40 PCR cycles (30 s at 95°C and 1 min at 60°C). An additional dissociation step of 15 s at 95°C, 20 s at 60°C and 15 s at 95°C was added to assess non-specific amplification. A negative control without cDNA template was included for each primer combination. Amplifications were performed in technical replicates consisting of independent cDNA syntheses derived from the same RNA sample and in three independent biological replicates. Ratios were calculated from the transcript numbers and normalized to *recA* as described in (Gonzalez-Rizzo et al., 2006).

Identification of ortholog of EcxR in *E. ruminantium* genome

To identify homologs of *Ehrlichia* EcxR in the genome of *E. ruminantium*, we used the same strategy as previously described in (Li and Carlow, 2012). The protein sequence of ECH_0795 (YP_507593) was used as query to search the genome of *E. ruminantium*. Multiple sequence alignment was done using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Larkin et al., 2007). Sequence identity values between the two sequences were generated using BlastP.

Cloning and expression of *erxR*

Full-length *erxR* was PCR amplified, using the primers shown in Table 2, and ligated into the NdeI and XhoI sites of the pET29a(+) vector (Novagen). The resulting plasmid was cloned into *E. coli* DH5α (Invitrogen) for amplification. The resulting plasmid, pErxR, was then purified using the QIAGEN Plasmid Maxi Kit (Qiagen, France) and cloned into *E. coli* BL21 (DE3) (Invitrogen) for protein expression. Expression of the protein was induced with 4 mM isopro-pyl-β-D-thiogalactopyranoside (IPTG) in 250 ml terrific broth. The protein then was purified

using Ni-NTA Fast Start Kit (Qiagen, France). ErxR expression was determined by Western blot analysis using anti-His tag antibody (Qiagen, France).

Construction of pUA66-derived promoter plasmids

The pUA66 plasmid was used for the analysis of promoter activity, using expression of a Green Fluorescent Protein (GFP). The promoters were PCR amplified from the genomic DNA of *Ehrlichia ruminantium* Gardel strain, using the primers shown in Table 2. Forward and reverse primers contained HindIII and BamHI restriction sites for cloning into the pET29a(+) plasmid. After cloning in pET29a(+), the promoters were digested with XhoI and BamHI for directed cloning into the pUA66 plasmid (Castaño-Cerezo et al., 2011). BL21 (DE3) cells were cotransformed with pErxR and each of the GFP reporter constructs, individually. The pET29a(+) vector alone and pUA66 containing the promoters but without co-transformation were used as a negative control. Cotransformants were grown in LB medium supplemented with 50 µg/ml kanamycin at 37°C for 2 h, followed by induction with 1 mM IPTG for 4 h. Induced bacteria were visualized as described below.

Microscopy

A drop (6 µl) of *E. coli* BL21(DE3) co-transformed with the pUA66 promoter and the pET29a-erxR plasmid (Table 3) suspended in LB growth medium was spotted into Superfrost plus slides (Fisher Scientific Ltd, UK) and visualized using an epifluorescent microscope Nikon Eclipse 80i (Nikon, France). Fluorescent images were acquired with a Nikon digital camera DXM1200F (Nikon, France), using Nikon ACT-1 software (Nikon, France). Fluorescence intensity was calculated by measuring the area, integrated intensity and mean gray value of the fluorescent bacteria and the background with ImageJ (National Institute of Health, USA). Corrected total cell fluorescence (CTCF) was calculated using the following formula: integrated density – (area of the cell x mean background readings). The average and statistical differences between the bacteria containing the plasmids with the different promoters and controls were calculated using the CTCF values from four different fields of view. Images were processed to size and adjusted in brightness and contrast using Adobe photoshop cs5 (Adobe Systems Inc., California, USA).

Statistical analyses

Statistical analyses were performed using Student's *t* test and a *P* value of < 0.05 was considered significant.

Neutral red staining

To investigate autophagic vacuoles, neutral red staining was used. Briefly, the starved cells of medium were stained with 0.05% neutral red at 28 °C for 2 min, and washed twice with PBS, then observed by optical microscopy.

Results

Identification of one EcxR homolog in *E. ruminantium* genome

We used the EcxR sequence (YP_507593.1) from *E. chaffeensis* to search NCBI databases, using the BLAST tool. Using this approach, we identified ERGA_CDS_03000 (YP_196226.1) as the closest homolog to EcxR in *E. ruminantium.* The results from the BLAST research revealed a putative conserved domain belonging to the HXT_XRE superfamily of DNA binding proteins (cl17200). Helix-turn-helix structure was defined according to the structure found in *Wolbachia* (Larkin et al., 2007). This protein of 124 amino acids has a predicted molecular mass of 14.25 kDa. Alignment of the deduced amino acid sequences of the various homologs is shown in figure 1A, including the transcription factor ApxR of *Anaplasma phagocytophilum*. By analogy to ApxR and EcxR, we named this protein *E. ruminantium* expression regulator, ErxR. The comparison of sequence identities of these proteins revealed a high degree of conservation (82% identity) between EcxR and ErxR. ApxR and ErxR showed 40% identity at the amino acid level. Structural analyses indicated that all orthologs shared a conserved helix-turn-helix domain that may function as a sequence specific DNA binding domain, such as in transcription regulators (Aravind et al., 2005).

Architecture of T4SS and *map1* gene cluster of *E. ruminantium*

We compared the genetic arrangement of *E. ruminantium* to that of *E. chaffeensis.* The five *virBD* loci are represented in figure 1B and 1C. In *E. ruminantium*, the genome sequence revealed the presence of two operons. *virD4*, *virB11*, *virB10*, *virB9a*, and *virB8a* are localized in operon 1. Operon 2 is located in the negative

A

```
EcxR            ----------------MTTISNQNDYGTKYKILKLIKDITYKNHWNQVTTARILGVDQPK
ApxR            ----------------MREKRRRNGQKIKLQILNIIRDTIQKSQWSQMAAAKVIGVDQPK
ErxR            MIDQGYFVFANLWYKIMTTISKTNDYTIKYRILKLIKDITYNNHWNQVTSAKILGVDQPK
                                *    . *.   * :**::*:*   :.:*.*:::*:::******
                                                helix            helix    helix

EcxR            ISQISNGKTAGFSLERLLIFLLRLKCDVNITITVNNPEIALINKDNMAESS--LESINLV
ApxR            VSQIINGKASGFSLERLLVFLLRLHCKVELSVSIGEIPQNILNEVEGSQLSDDLSAIDLK
ErxR            ISHISNGKTAGFSLERLLIFLLRLKCDINITITVNNSEIALINKDSISDKS--LESINLV
                :*:* ***::********:*****:*.:::::::.:    ::*: . :: *  *.:*:*
                helix         helix     β-strand

EcxR            IIQKED-
ApxR            VISKDSK
ErxR            ILQKNN-
                ::.*:.
```

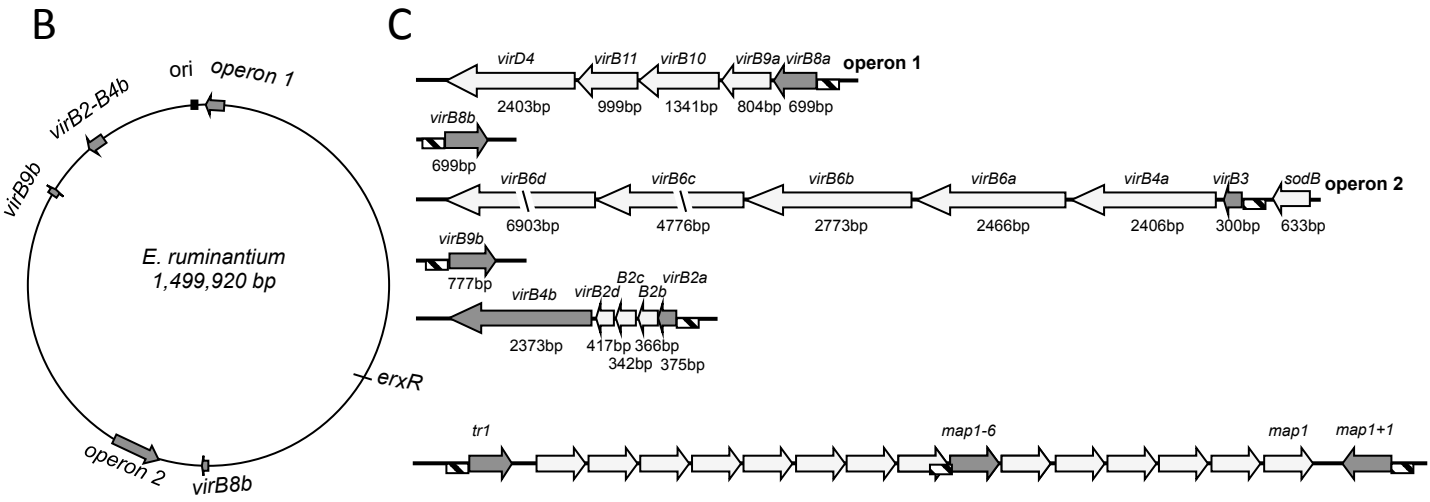


Figure 1

strand and contains four copies of *virB6*, along with one copy of *virB4a* and *virB3,* all located downstream *sodB*. Four duplicated versions of *virB2* are found scattered through the genome. Likewise, we present the arrangement of the *map1* family previously reported by (Postigo et al., 2007) (figure 1C).

Analysis of *erxR* expression during the life cycle of *E. ruminantium*

To determine the relative expression of *erxR* throughout the developmental cycle of *E. ruminantium*, *erxR* mRNA expression was analyzed by qRT-PCR. The relative expression of *erxR* was calculated by dividing the transcripts number to the total number of bacteria at each time point. Fold change was then calculated by comparing the relative expression from each time point to the relative expression at 96 hpi, the stationary phase. The expression of *erxR* increased by 0.5 fold at 24 hpi, decreased in expression at 48 and 72 hpi, and peaked at 120 hpi increasing by 4 fold (figure 2), which corresponds to the time of lysis. The results suggest that the expression of *erxR* occurs before bacteria enter to the host cell and at early stages of development cycle *in vitro*.

rErxR activates *gfp* reporter fusions

To examine if ErxR can activate the expression of the *virBD* genes, *map1+1*, *map1-6*, *tr1* and *erxR*, the promoters for these genes were cloned into a *gfp* reporter plasmid and transformed into *E. coli* BL21 (DE3) carrying pErxR or empty pET29a(+) vector. The *virB3-gfp* reporter constructs presented a significant increase in fluorescence intensity after IPTG induction (27000 units of fluorescence intensity) compared to samples lacking IPTG (1000 units of fluorescence) or compared to the control (5000 units of fluorescence (figure 3B). Activation was also observed for *tr1* and *erxR* promoter (figure 3A, 3C). No activation was observed for *virB9* promoter (data not shown). Western blotting experiments confirmed expression of ErxR was only detected following induction with IPTG (figure 4).

Evaluation of the pH inside *E. ruminantium* vacuoles by neutral red

Recently, it has been shown that *E. chaffeensis* vacuoles are acidified with an approximate pH of 5.2 (Y. Cheng et al., 2014). To determine if *E. ruminantium* vacuoles were acidified as well, we stained the morulae with neutral red that is normally used to stain lysosomes. Microscopy demonstrated that *E. ruminantium*

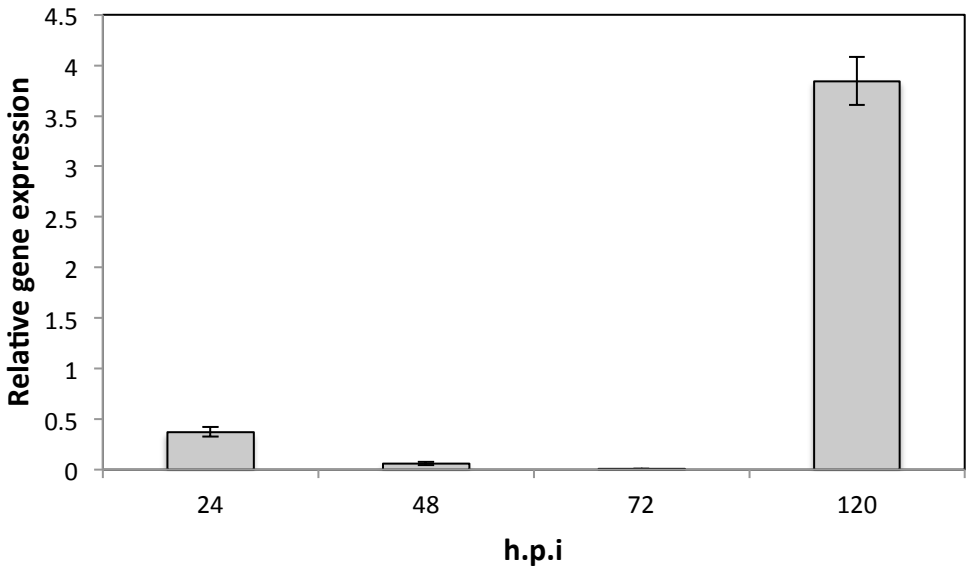


Figure 2

vacuoles stained in red, suggesting that *E. ruminantium* vacuoles are also acidified (figure 1 supplementary data).

Acidity induces the expression of some T4SS and *map1* genes in *E. ruminantium*

In order to determine if the acidification of the *Ehrlichia* morulae plays a role in regulating the T4SS activity, we studied the expression of T4SS operons during *in vitro* culture of *E. ruminantium* under acidic conditions by qRT-PCR. *E. ruminantium* was incubated in medium BHK21 at pH 7.0 (neutral) or 6.0 (acidic) during 24 h after lysis and then RNA was extracted for analysis. This time was chosen because it simulates the early stages of development, i.e. at 24 hpi, which corresponds to the time when the bacteria are in the reticulate form. Each *virB* gene upstream the five *virBD* loci were chosen for expression analysis (figure 5A). Four genes of the T4SS, *virB2a*, *virB8b*, *virB3* and *virB4a,* presented significant differences in expression under acidic pH with changes of 1.5, 1.5, 0.8, and 2.5 fold up-regulation, respectively. Expression of *virB4b*, *virB9b*, and *virB8a* (the gene upstream of operon 1) did not change significantly between acidic and neutral pH. The *resolvase* gene was used as control as it is not known to respond to pH change. Because our results showed that ErxR regulated some of the genes in the T4SS of *E. ruminantium*, we examined if low pH had an effect on the expression of *erxR* (figure 5B). Real-time PCR assays indicated that the *erxR* expression is up-regulated by 1.6 fold at acidic pH.

Additionally, we tested whether or not the expression of the members of the *map* family *map1-6* and *map1+1*, localized in the middle and the border of the *map1* cluster respectively, changes under low pH conditions (figure 5C). According to our results, only *map1+1* is up-regulated in response to acidification by 2 fold, whereas *map1-6* was not significantly up-regulated.

The expression of T4SS and *map1* genes is induced by iron depletion in *E. ruminantium*

Iron is essential for various bacterial biophysical processes. Iron uptake mechanisms are closely associated with bacterial pathogenesis and may be connected to the expression of certain virulence determinants in *E. ruminantium.* Therefore, we studied the T4SS expression in response to iron starvation. We incubated the bacteria in media containing iron or an iron chelator (BPD) for 24 h after lysis as

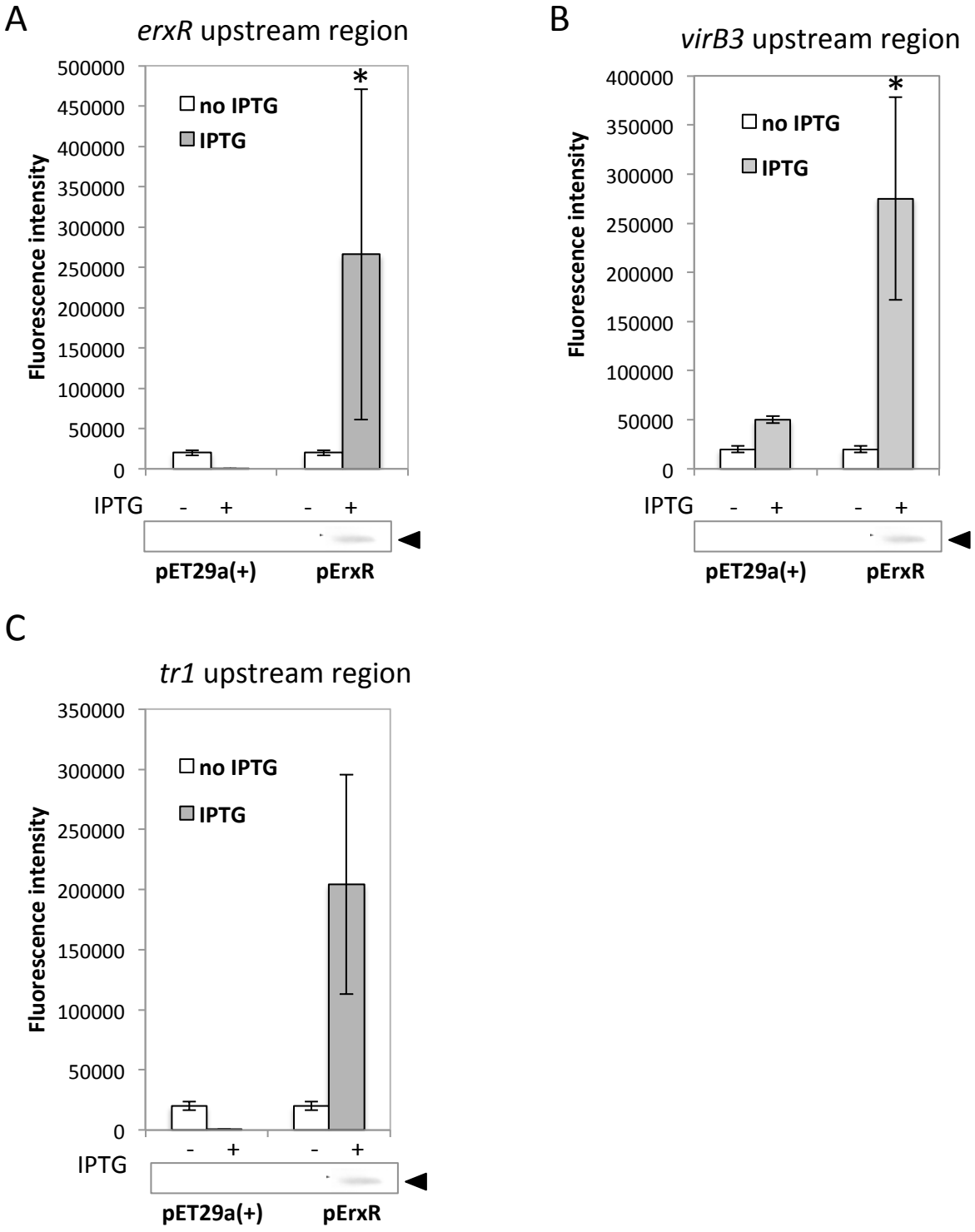

A
erxR upstream region
Fluorescence intensity
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
no IPTG
IPTG
*
IPTG - + - +
pET29a(+)
pErxR
B
virB3 upstream region
Fluorescence intensity
400000
350000
300000
250000
200000
150000
100000
50000
0
no IPTG
IPTG
*
IPTG - + - +
pET29a(+)
pErxR
C
tr1 upstream region
Fluorescence intensity
350000
300000
250000
200000
150000
100000
50000
0
no IPTG
IPTG
- + - +
IPTG
pET29a(+)
pErxR

Figure 3

previously described for the acidic conditions experiment. Six components of the T4SS, *virB3*, *virB4a*, *virB4b*, *virB8a*, *virB8b*, *virB9b* were significantly up-regulated under iron-depletion (10 to 190 fold increase) (figure 6A). *virB2a* was the only gene that did not present a change in expression (figure 6A). The *erxR* gene was up-regulated during iron starvation by 25 fold (figure 6B).

We suspected that some of the Map1 homologs might function as TBDRs, which are OMPs known to be involved in the acquisition of iron under iron starvation. Therefore, we tested the expression of two *map1* genes, *map1+1* and *map1-6*, under iron starvation. Culture under iron limitation strongly increased the expression of both genes (figure 6C). The expression level of *map1+1* was 400 fold higher under iron-depleted condition (figure 6C), suggesting that this gene may play a role during iron starvation.

Because of the strong effects of iron depletion in the expression of the T4SS genes and *map1* genes, we used bioinformatic analysis to determine the presence of a Fur box close to the *erxR* promoter region 400 bp upstream of the initiation codon. Fur boxes work as repressors when iron is present and thus are involved in the regulation of genes for iron acquisition. Using custom program based on regular expression searches, we determined the presence of a Fur box in the position -35 to -10 upstream of the *erxR* initiation codon (figure 2 supplementary data). The alignment of the Fur box sequence with that of *E. coli* showed the presence of 2 nATWAT motifs separated by 6 nucleotides with only 5 mismatches between the Fur box of *E. ruminantium* and *E. coli* (figure 2 supplementary data). The presence of a Fur box and the similitude with other known Fur box, suggests that iron concentrations in the media may affect the expression of *erxR*, which may affect other genes regulated by this transcription factor as suggested by our results.

Discussion

The preferential expression of virulence factors with diverse functions as to host clues and the environment has been previously characterized in several bacteria (Oogai et al., 2011; Weber et al., 2014). Likewise, several *Ehrlichia* and other members of the *Anaplasmataceae* family have been shown to differentially express certain genes in response to the host cells (Nelson et al., 2008). Among these genes are those encoding for components of the T4SS, which have been shown to play an

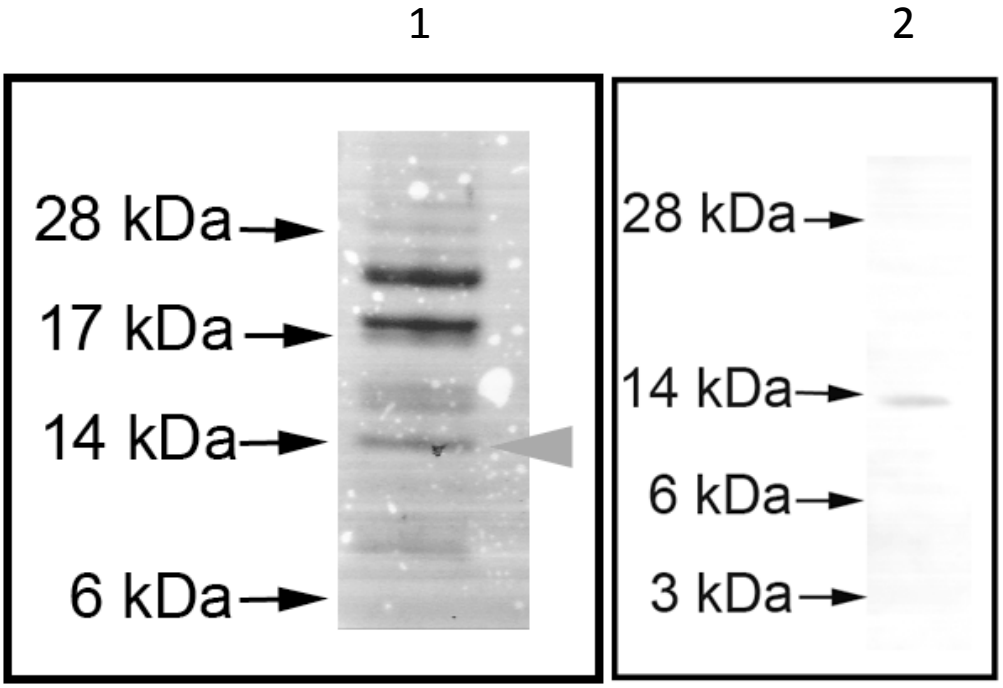


Figure 4

essential role in pathogenicity. Likewise, OMPs are often regulated by environmental signals and play an important role in bacterial pathogenesis by enhancing the ability of the bacteria to adapt to different environments. Interestingly, T4SS affects OM properties that might be important for the adaptation of *Brucella* to both *in vitro* and *in vivo* (Y. Wang et al., 2010). Herein, we present the effects in *E. ruminantium* of two environmental clues, pH and iron concentration, on the expression of the transcriptional regulator *erxR*, the components of the T4SS apparatus and members of the MAP1 family. Our results indicate that iron concentration and pH could be used by *E. ruminantium* as clues for intracellular growth and lead to the differential expression of the genes mentioned above.

Many intracellular bacterial pathogens use T4SS to deliver effector molecules that subvert the eukaryotic host cell defenses and other processes in their advantage (Trokter et al., 2014). The genetic arrangement of *E. ruminantium* T4SS genes is similar to that of *E. chaffeensis*. T4SS is composed of 12 proteins that assemble in the envelope of the bacterium (figure 1). The regulation of the components of the T4SS, the *virBD* genes, is affected by their function. In *E. chaffeensis*, the *virBD* genes were up-regulated during the early exponential phase. The expression of the two major operons of the T4SS and the three duplicated genes peaked at 24 hpi and 48 hpi, respectively. This corresponds to early time-points before exponential growth when the infectious elementary bodies differentiate in replicative reticulate bodies (Z. Cheng et al., 2008). Although we did not studied the expression of T4SS components during the developmental cycle of *E. ruminantium*, it is very likely that it is similar to that of *E. chaffeensis* and that the regulation of these components is driven by the same regulatory protein.

Regulatory proteins are known to play an important role in the survival and persistence of intracellular pathogens in their host (Zusman et al., 2007; De Jong et al., 2008; Altman and Segal, 2008; Martínez-Núñez et al., 2010). EcxR (ECH_0795) is the only transcriptional regulator that has been associated with the expression of T4SS components (Z. Cheng et al., 2008). We identified one ortholog (*erxR*) in *E. ruminantium* (figure 1A), which binds and regulates certain promoters of the *virBD* genes, the *tr1* promoter upstream of the *map1* genes, and itself (figure 3). Using qRT-PCR at different time points, we determined that *erxR* is strongly expressed at

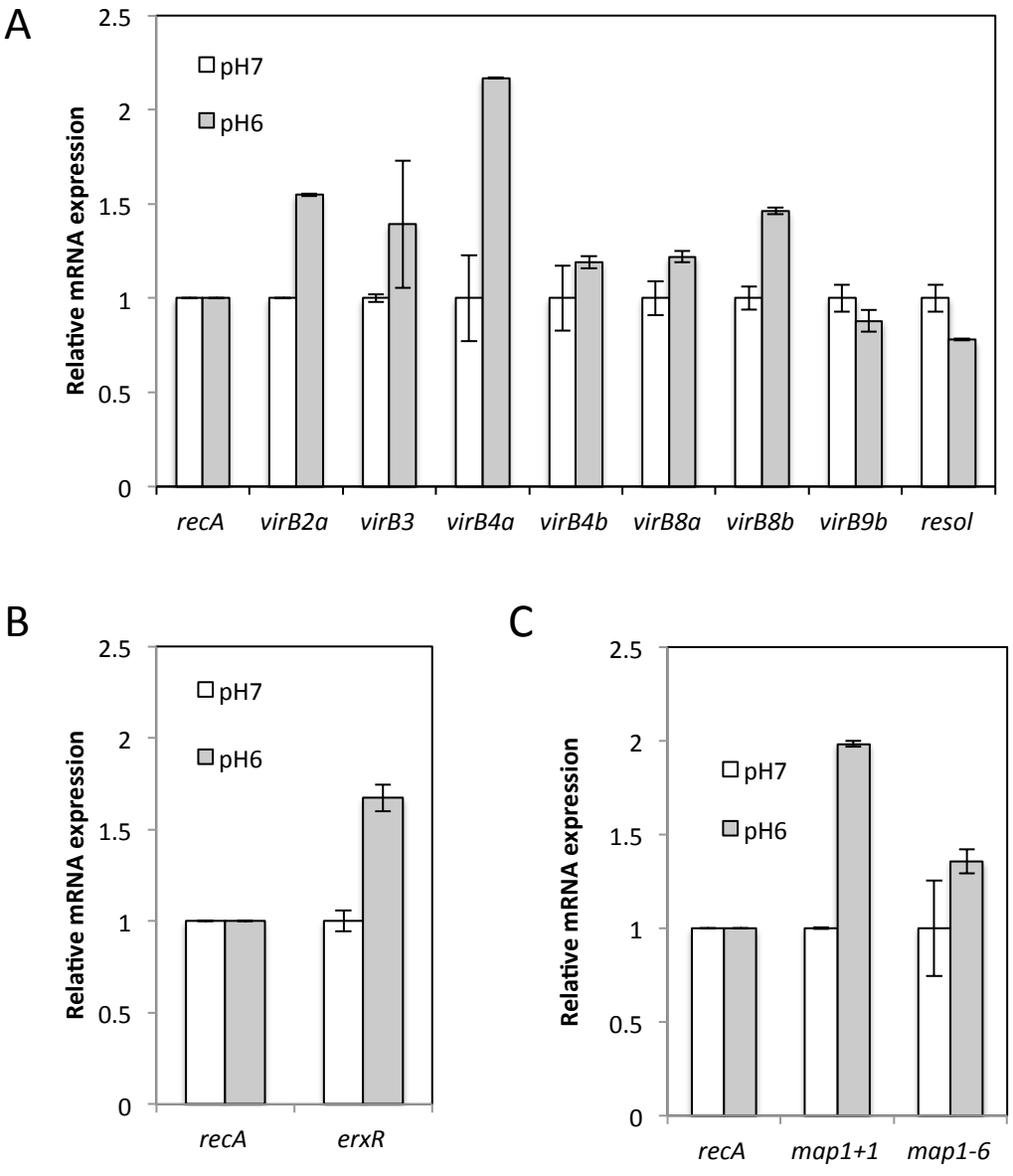

A
2.5
2
1.5
1
0.5
0
pH7
pH6
Relative mRNA expression
recA
virB2a
virB3
virB4a
virB4b
virB8a
virB8b
virB9b
resol
B
2.5
2
1.5
1
0.5
0
pH7
pH6
Relative mRNA expression
recA
erxR
C
2.5
2
1.5
1
0.5
0
pH7
pH6
Relative mRNA expression
recA
map1+1
map1-6

Figure 5

elementary bodies stage (120 hpi and during early developmental stages of *E. ruminantium* (24 hpi). If the expression of the T4SS components in *E. ruminantium* is similar to the of *E. chaffensis*, as we suspect, the results suggest that the regulator *erxR* is expressed at the same time as the T4SS during developmental cycle and that plays an important role in the regulation of the T4SS.

erxR and its homologs in other *Anaplasmataceae* appear to also be associated with the expression of important antigenic OMPs. ApxR, one homolog of ErxR, regulates the transcription of *p44* transcription by binding to the *tr1* promoter during *A. phagocytophilum* infection of mammalian host cells (X. Wang, Cheng, et al., 2007; X. Wang, Kikuchi, et al., 2007). *p44E* encodes the immunodominant pleomorphic 44-kDa major surface protein that shows homology with the MAP1 family in *E. ruminantium* as well as the p30 family in *E. canis* and the p28 family in *E. chaffeensis* (Dunning Hotopp et al., 2006). The major antigenic protein MAP1 is part of a multigene family containing 16 paralogs tandemly organized in a head to tail arrangement that are downstream of a hypothetical transcriptional regulator gene (*tr1*) (Postigo et al., 2007) (figure 1C), a similar arrangement to that reported for *p44, p30,* and *p28* in the other *Anaplasmataceae* (Dunning Hotopp et al., 2006). *tr1* is one of the three promoters that have been identified in the *p44* expression locus and it has been shown to be the strongest promoter that drives the expression of a polycistronic mRNA containing OMP1, *p44ESup,* and *p44* (Barbet et al., 2005). *tr1* harbors a winged helix-turn-helix and a DNA binding motif and its part of the xenobiotic response element family of transcriptional regulators. However, the function of *tr1* remains unclear (Nelson et al., 2008), and whether or not *tr1* drives the expression of a polycistronic tandem mRNA containing several *map* homologs is still unknown. According to our results, ErxR binds to *tr1* (figure 3C) and it is possible that it regulates the expression of the *map1* members in a similar manner as *tr1* in *A. phagocytophilum* for *p44* expression. Interestingly, in the endosymbiotic bacterium *Wolbachia* (wBm) wBmxR1 and wBmxR2, homologs to ErxR, were shown to co-regulate genes of the T4SS and riboflavin biosynthesis pathway (Li and Carlow, 2012). Riboflavin is a co-factor important for the survival of the endosymbiont's host, the filarial parasite *Brugia malayi* (Li and Carlow, 2012). Thus, this family of regulators appears to activate the expression of genes with variable functions that

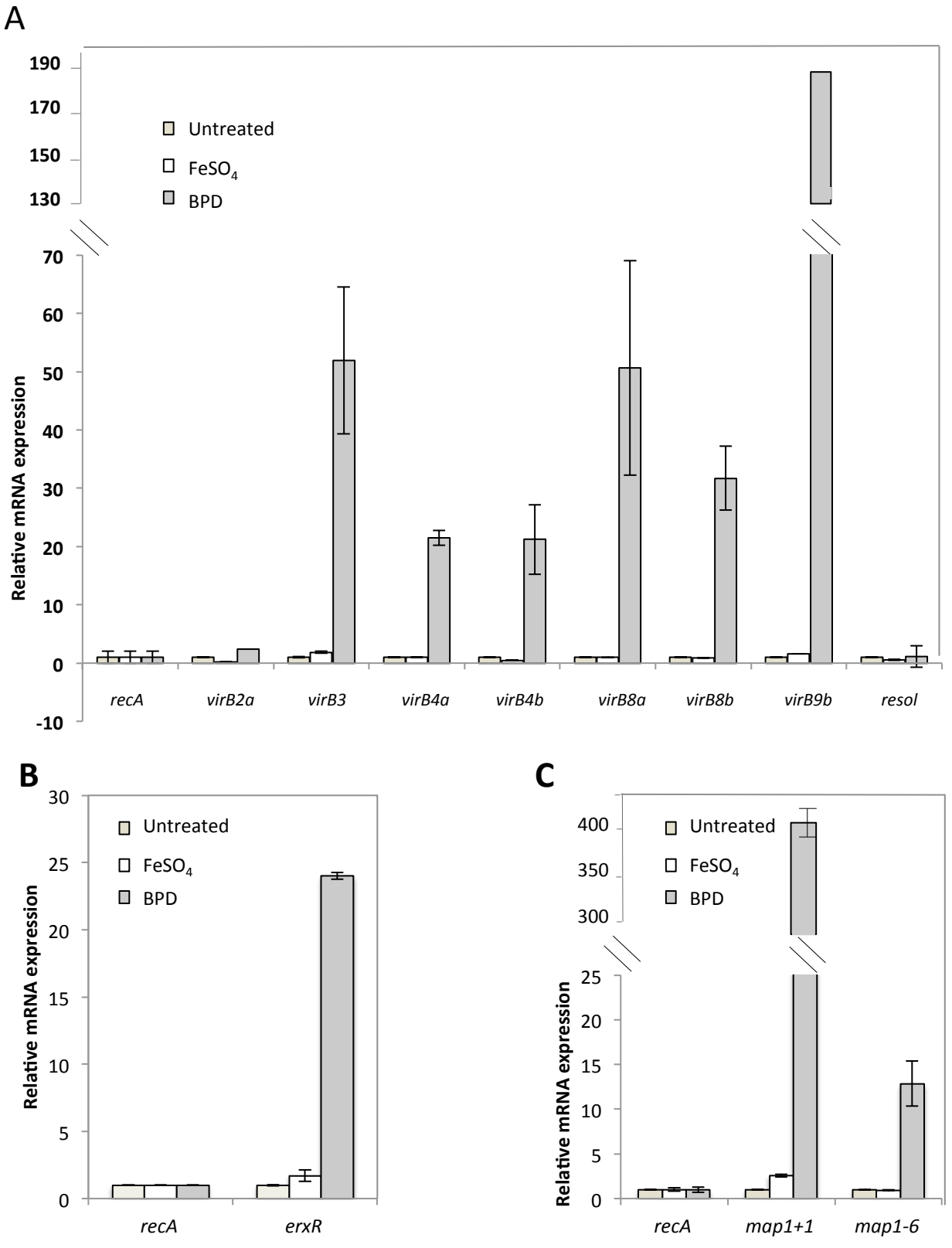

A
190
170
150
130
70
60
50
40
30
20
10
0
-10
Untreated
FeSO4
BPD
Relative mRNA expression
recA
virB2a
virB3
virB4a
virB4b
virB8a
virB8b
virB9b
resol
B
30
25
20
15
10
5
0
Untreated
FeSO4
BPD
Relative mRNA expression
recA
erxR
C
400
350
300
25
20
15
10
5
0
Untreated
FeSO4
BPD
Relative mRNA expression
recA
map1+1
map1-6

Figure 6

are involved in the pathogenicity and intracellular survival of these bacteria.

The regulation of the T4SS and members of the MAP1 by ErxR might be triggered by environmental and nutritional clues in the host cell. Our transcriptional analysis shows that the expression of the regulatory gene *erxR*, *virB* genes and *map1* genes changes in response to two important environmental signals: acidic pH and iron starvation (figure 5 and figure 6). It is possible that the regulator ErxR is activated by an unidentified sensor kinase that responds to these clues and up-regulates expression of the T4SS genes and certain *map1* genes in response to low pH and iron starvation.

A recent study revealed that the *E. chaffeensis* vacuole in DH82 cells is acidified at pH 5.2 (Y. Cheng et al., 2014), suggesting that *E. ruminantium* also resides in an acidified vacuole. Acidification of *E. ruminantium*'s morulae in BAE infected cells was determined using red neutral at 96 hpi. Low pH is one of the signals that trigger expression of a virulence factors that are essential for the biogenesis of the intracellular compartment (Porte et al., 1999). Thus, the acidification of *Ehrlichia*-containing vacuole may be necessary for intracellular survival and replication of *E. ruminantium*. Phagocytosis is an important part of host defenses, playing a critical role in the innate immune response against pathogens and in the initiation of adaptive immunity. However, numerous pathogens have evolved complex mechanisms to manipulate the phagocytic vacuole to establish a survival niche by injecting proteins that modify it (Ray et al., 2009). This process starts with the induction of the secretion system by the acidification of the vacuole. For example, acidification is an important signal for activation of *Salmonella* spp. T3SS, which is required for the intracellular replication and survival of the bacteria inside phagocytes (Rappl et al., 2003; Arpaia et al., 2011). It has been demonstrated that the activation of SsaL, a transcriptional regulator, by acidic conditions is required for the synthesis of the T3SS apparatus and the translocation of T3SS effectors (Coombes et al., 2004). Likewise, a similar induction of the transport system by acidic conditions and nutritional clues has been reported in *Edwardsiella* species and *Agrobacterium tumefaciens* (Chakraborty et al., 2010; Rogge and Thune, 2011; Wu et al., 2012; Baumgartner et al., 2014), *Edwardsiella* species's T3SS is activated by low pH and limited phosphate through the control of the EsrA/B two-component regulatory

system (Rogge and Thune, 2011), which is similar to our finding that both acidification of the vacuole and iron starvation lead to the expression of *E. ruminantium*'s T4SS. Interestingly in the case of *Salmonella*, both translocators and effectors are only expressed in acidic minimal medium and not in minimal medium at neutral pH. The conditions in the acidic minimal medium mimic those of the vacuole (Yu et al., 2010). Whether or not this is the case in *E. ruminantium* is not known. However, it would be interesting to see if among the putative T4SS effectors identified by the S4TE algorithm in *E. ruminantium* (Meyer et al., 2013), one is expressed only under low pH.

In the same manner, our results suggest that ErxR activates the expression of T4SS genes and certain *map1* genes under iron starvation. Acquisition of iron is often subject to competition between bacteria and the host. Many bacterial pathogens sense iron depletion as signal that they are within a vertebrate host (Skaar, 2010). This sensing involves transcriptional control mediated by the transcriptional repressor, Fur (Escolar et al., 1998). But many organisms, including *E. coli*, *Campylobacter jejuni*, *Vibrio cholerae*, and *Vibrio vulnificus*, have been shown to utilize Fur to positively regulate gene expression with increasing iron concentrations (O'Sullivan et al., 1994; Palyada et al., 2004; Mey et al., 2005). For example, Fur activates *sodB*, an iron superoxide dismutase, under iron dependent conditions. SODs are metalloproteins that play an important role in protection against oxidative stress by catalyzing dismutation of superoxide radical (O_2^-). Interestingly, *sodB* in *E. ruminantium* is located upstream of operon 2 of the T4SS and it is co-transcribed along these genes in *Ehrlichia* and *Anaplasma*, suggesting an effect of iron in the expression of *virBD* genes. Moreover, it has been demonstrated that the regulator EcxR binds to the promoter regions upstream of *sodB.* In our model, Fur free from ferrous iron can activate *erxR* under low iron conditions. We searched for Fur boxes in the genome of *E. ruminantium* and found one upstream of *erxR*. The 19 bp sequence consisted of two repeated hexamers (nATWAT) flanking a 7 nt sequence, which is commonly found upstream iron regulated enzymes such as succinate dehydrogenase iron-sulfur subunit, major ferric iron binding protein precursor, and adenosine tRNA methylthiotransferase (Escolar et al., 1998; Grifantini et al., 2003). Thus, it is possible that *E. ruminantium* is capable of sensing low iron concentrations

in the environment and activates the expression of *ErxR* through this Fur box.

However, how *Ehrlichia* senses low iron concentrations is not known. It is possible that some Map1 proteins may play a similar role as TBDRs in the perception of environmental cues and in iron uptake (Blanvillain et al., 2007). As mentioned above, Map1 proteins are homologs of members of the *p28* family in *E. chaffeensis*, which have been shown to function as porins and possibly act in nutrient uptake during intracellular infection (Kumagai et al., 2008). It has been shown that porins, such as OmpA and OmpC, bind to transferrin and act in iron uptake of the enteropathogenic strains of *E. coli, Salmonella typhimirum,* and several *Shigella* species (Sandrini et al., 2013). Similarly, *Microbacterium smegmati* is able to acquire ferric ions under low-iron conditions through members of the Msp family of porins (Jones and Niederweis, 2010). The up-regulation of *map1+1* and *map1-6* under iron starvation is an indication that these two porins may play a role in iron acquisition and rises the possibility that MAP1 proteins may also act as sensor, although more evidence is needed to conclude. These results suggest that the *map1* genes may fulfill several functions during infection and it will be interesting to characterize these functions in order to better understand the adaptation of this pathogen to its host as done by Blanvillain *et al.* (Blanvillain et al., 2007).

In conclusion, we have demonstrated that exposure of *E. ruminantium* to acidic pH or iron limitation induces ErxR-dependent expression of T4SS apparatus and *map1* genes. These findings reveal an important degree of coordination between T4SS and *map1* genes at the transcriptional level and contribute to a better understanding of the infection process. Final analysis of the data presented here enabled us to develop a model for the regulation of *E. ruminantium*'s T4SS and *map1* genes (figure 7).

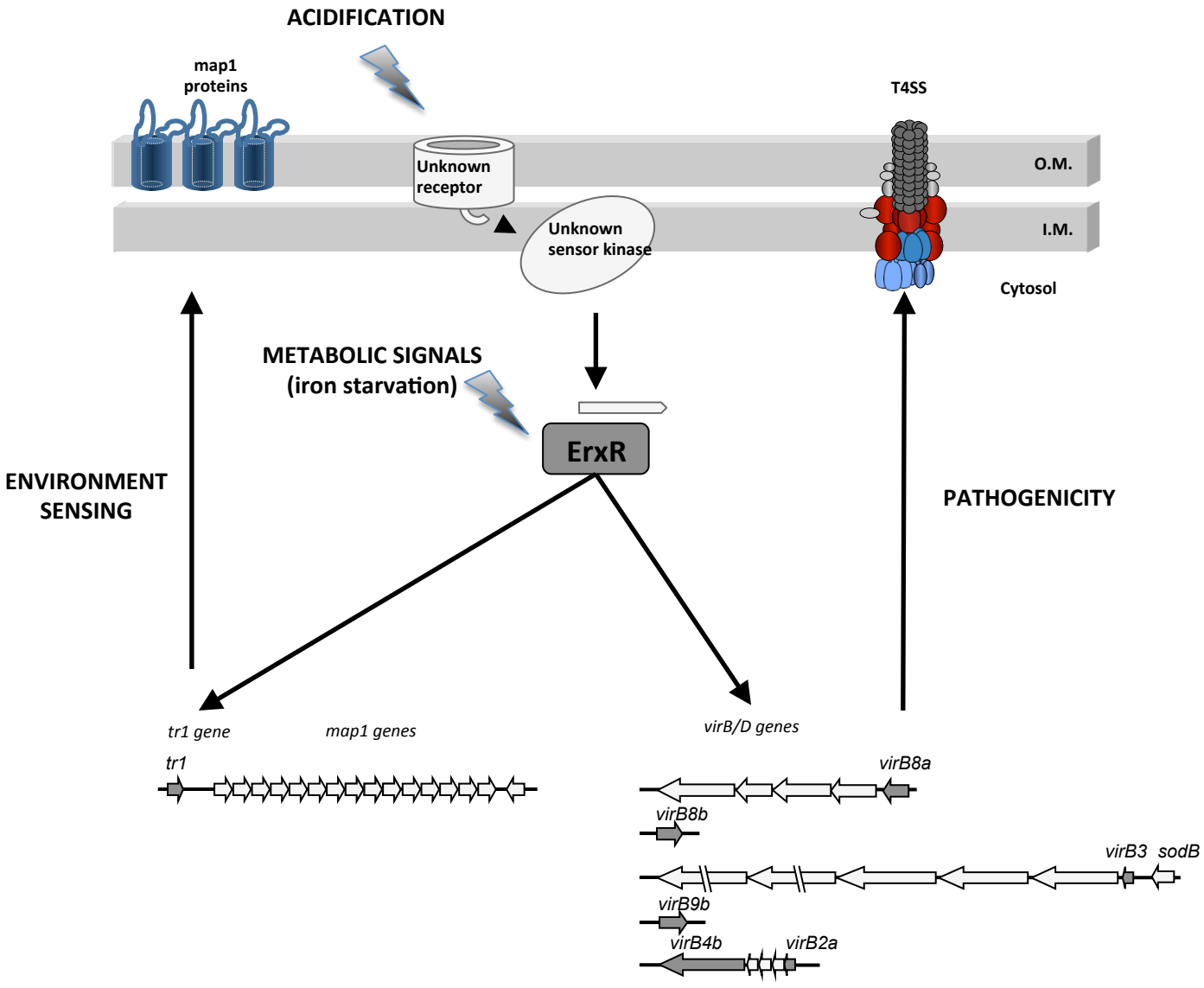

ACIDIFICATION
map1
proteins
T4SS
Unknown
receptor
O.M.
I.M.
Unknown
sensor kinase
Cytosol
METABOLIC SIGNALS
(iron starvation)
ErxR
ENVIRONMENT
SENSING
PATHOGENICITY
tr1 gene
map1 genes
tr1
virB/D genes
virB8a
virB8b
virB3
sodB
virB9b
virB4b
virB2a

Figure 7

Acknowledgments

We gratefully thank Dr. Sara Castaño-Cerezo for providing pUA66 plasmid.
The authors acknowledge financial support from FEDER grant FED 1/1.4-30305, 2007–2013, “Risque en santé animale et végétale” (PhD grant to AM).

Table and Figure legends

Figure. 1. Potential homolog of the *Ehrlichia* type IV secretion system regulator EcxR in *E. ruminantium* and organization of *virBD* loci and *map1* cluster . (A) Amino acid alignment of EcxR and its homologs from *E. ruminantium* (ERGA_CDS_03000) and *A. phagocytophilum* (ApxR). The alignment was generated using ClustalW (http:/ www.ebi.ac.uk/Tools/msa/clustalw2/). Identical and conserved residues are indicated with stars and points, respectively. The positions of amino acids predicted to form a helix structure and beta-strand are shown. (B) The genomic map of *E. ruminantium* is represented as a circle. The origin of replication (ori) is indicated by a black box. Gray arrows show the *virBD* gene loci and cluster *map1*. T4SS operon 1 consists in *virB8a, virB9a, virB10, virB11* and *virD4* and operon 2 consists in *sodB*, *virB3*, *virB4a*, *virB6a*, *virB6b*, *virB6c* and *virB6d*. Four copies of *virB2* have been identified upstream of *virB4b*. *map1* genes belongs to a multigenic family organized in one cluster of 16 paralogs localized downstream the transcriptional factor *tr1*. Length of arrows is proportional to the length of the gene. (C) Cluster structure of *virBD* gene loci and *map1* genes. *virBD* and *map1* genes are represented by gray arrows. The names and length of the genes are indicated above and below the arrows. The upstream regions amplified by QRT-PCR for expression assays (hatched boxes) are indicated.

Figure. 2. *erxR* is expressed early during infection and peaks at 120 hpi in *E. ruminantium*. Quantitative RT-PCR was used to determine the temporal expression of *erxR*. Relative expression at different developmental stages was normalized by dividing to the number of bacteria. Fold differences were evaluated by comparing each time point to 96 hpi (the stationary phase). Data were obtained from triplicate samples and expressed as a mean $\pm$ standard deviation.

Figure. 3. ErxR activates the transcription of *virBD gfp* reporter fusions. Fluorescence intensity was used to measure the transcriptional activities of *gfp* reporter constructs. The values are the means ± standard deviations for three specimens. An asterisk indicates the that value are significantly different ($P < 0.001$) from the controls. Western blot analyses were performed in samples from the fluorescence assays using an anti-His tag antibody to verify the expression of rErxR. The position of rErxR is indicated by arrowheads.

Figure. 4. Production of rErxR in *E. coli*. *E. ruminantium erxR* was cloned into the pET29(+) vector, expressed, and purified using nickel chelate chromatography. The purified protein was subjected to SDS- PAGE analysis, followed by Coomassie blue staining (lane 1) and Western blot analysis using an anti-His tag antibody (lane 2). Lane M contained pre-stained protein size standards. Each lane contained 8 μg of recombinant protein.

Figure. 5. T4SS and MAP encoding genes as well as *erxR* are up-regulated at low pH. The expression of the *virB* (A), *erxR* (B) and *map1* (C) genes were measured at the lysis phase of infection under pH 6.0 or 7.0 using quantitative real-time PCR. The data represent the mean ± SD of 2 or 3 biological replicates, each of which contained 3 technical replicates.

Figure. 6. T4SS and MAP encoding genes as well as erxR are up-regulated under iron-depletion. The expression of the *virB* (A), *erxR* (B) and *map1* (C) genes were measured at the lysis phase of infection under iron-repletion or iron depletion using quantitative real-time PCR. The data represent the mean ± SD of 2 or 3 biological replicates, each of which contained 3 technical replicates.

Figure. 7. Putative regulation model from the T4SS and *map* genes in *E. ruminantium* under acidic and iron starvation conditions.

Supplementary Figure. 1. Acidification of of *E. ruminantium* morulae. Infected BAE cells at 96 hpi were incubated for 2 min with 0.05% neutral red and washed. The morulae stained in red dye, which indicates pH under 6. The arrowheads indicated the positioning of the morulae within infected cells

Supplementary Figure. 2. (A) Nucleotide sequence of predicted *erxR* promoter. The initiation codon of *erxR* is shown in capital letter at the end of the sequence. The

sequence in red represents the nATWAT motif. The -35 and -10 boxes of the *erxR* promoter are capitalized. (B) Potential *erxR* Fur box is compared with the consensus *E. coli* Fur box. Conserved residues are shown in red.

TABLE 1. Oligonucleotide primers used for quantitative RT-PCR

Gene	Primer		Target size (bp)
	Direction[a]	Sequence (5'-3')	
map1	F	CACTTGAAGGAATGCCAGTTTCTC	85
	R	CTTAGGATTTGTAGCATTGATTACTGACACT	
virB2a	F	ATTTCTTAGCGTCTGATATCCCCTGC	227
	R	CCAAGTCAACAACAAAATAATGATGATGC	
virB3	F	CTAAGAATAACGGCTCTTTTGATGAAGC	170
	R	TGTTTATGGCAGGTACTGTTAAAGCAGA	
virB4a	F	AGTTCATGAGTACTGTCTTTCCACCACC	233
	R	GAAATTTTGAATATGTGGTGCGTAAAGC	
virB4b	F	TCACAAAAAAAATACCCTTCTTGAGATGG	185
	R	GTGAGAATATTCCAGCTGTTGTGTCAAG	
virB8a	F	TTGCTCATTCATTTCGAGAGTTG	120
	R	CCTTATTTGTACATTTCCTGGACGT	
virB8b	F	ACCGAACGTGCTTGGTTGTCA	206
	R	AGGTATCGTGATAGGAGTGTTTCTTCACT	
virB9b	F	ATTTGTTTTGCAAAGCAGGAAGTACG	199
	R	CGGTTACCTTTTGGGATTAACTGCC	
erxR	F	AGTAAGCGTTCTAGAGAAAATCCTGCTG	227
	R	TGATCAAGGTTATTTTGTCTTTGCGAAT	
resol	F	TCAATGGTAATTGGGTTGCCACTTGAT	158
	R	TTTGCTACCCTAGTAGCCATAGCAGT	
recA	F	TTGAAAAAGCGTTTGGTCGTG	120
	R	GGGAAACCACCAATACCCAAT	
map1+1	F	ACTCCAAATTTAAGCTGATAA	119
	R	TGATGTTGCTCACGGAAATA	
map1-6	F	ATACACCAACATTCCAGAACA	120
	R	CAGGGATTTCTGCATCGA	

[a]F, forward; R, reverse

TABLE 2. Oligonucleotide primers used to construct plasmids

Gene or region	Primer Direction[a]	Primer Sequence[b]	Target size (bp)	Enzyme	Plasmid
erxR	F	TATACATATGATTGATCAAGGTTATTTTG	375	NdeI	pET29a(+)
	R	GTGCTCGAGATTATTTTTTTGTAATATGACG		XhoI	
erxR upstream region	F	GGGAAGCTTTGTACCATTGATAATCGAG	400	HindIII	pUA66[c]
	R	CCCGGATCCTCAAATTTAAAAAATAATACTTA		BamHI	
virB3 upstream region	F	GGGAAGCTTTTGGGTTTTGCATCTCAAAG	451	HindIII	pUA66
	R	CCCGGATCCCCTGCCATAAACACATAACC		BamHI	
virB2a upstream region	F	GGGAAGCTTGTAGAGCTGATATGTAACATG	490	HindIII	pUA66
	R	CCCGGATCCCCATTAAAACCCCCCATAATC		BamHI	
virB2d upstream region	F	GGGAAGCTTCCATTAATAATCATTCACTAC	272	HindIII	pUA66
	R	CCCGGATCCTGAATAGCAGTAACTTTAATTG		BamHI	
virB9a upstream region	F	GGGAAGCTTACATTCTATGTCAGATCCAGC	428	HindIII	pUA66
	R	CCCGGATCCTAACTTGCAACAATTCCATGC		BamHI	
virB8b upstream region	F	GGGAAGCTTGTGGTATCTTGTCTGATCAAG	429	HindIII	pUA66
	R	CCCGGATCCAGAAATAGTAACGATGTTCG		BamHI	
map1+1 upstream region	F	GGGAAGCTTGATCATAATATAGTGTAATAG	426	HindIII	pUA66
	R	CCCGGATCCTAATAACCAAAGGTAAACTAC		BamHI	
map1-6 upstream region	F	GGGAAGCTTGCCAGACCAAAACAATTACAC	416	HindIII	pUA66
	R	CCCGGATCCAAATGCTCCTATAAATATGTTAACTG		BamHI	
tr1 upstream region	F	GGGAAGCTTAATCTCCTCATAAAAACTAAC	234	HindIII	pUA66
	R	CCCGGATCCATCATTTTCCTATAATAAGGC		BamHI	

[a]F, forward; R, reverse

[b]Enzyme sites are underlined

[c]See reference Castano-Cerezo, 2011

TABLE 3. Strains and plasmids used in this work

Strain	Properties	Source
Escherichia coli DH5α	F- ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rk$^-$, mk$^+$) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ-	Invitrogen
Escherichia coli BL21	F-*ompT hsdSB* (rB-mB-) *gal dcm rne131* (DE3)	Invitrogen
Plasmids		
pET29(+)	Expression vector C-terminal His6x tag (Km^r)	Novagen
pUA66	Promoter-probe plasmid using EGFP as a reporter of expression (Km^r)	Castaño-Cerezo *et al.*, (2011)

Km, kanamycin resistance

References

Abromaitis, S., Nelson, C.S., Previte, D., Yoon, K.S., Clark, J.M., DeRisi, J.L., and Koehler, J.E. (2013) *Bartonella quintana* deploys host and vector temperature-specific transcriptomes. *PLoS ONE* **8**: e58773.

Allsopp, B.A. (2010) Natural history of *Ehrlichia ruminantium*. *Veterinary Parasitology* **167**: 123–135.

Altman, E. and Segal, G. (2008) The response regulator CpxR directly regulates expression of several *Legionella pneumophila icm/dot* components as well as new translocated substrates. *Journal of Bacteriology* **190**: 1985–1996.

Andrews, S.C., Robinson, A.K., and Rodríguez-Quiñones, F. (2003) Bacterial iron homeostasis. *FEMS Microbiol Rev* **27**: 215–237.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., and Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**: 231–262.

Arpaia, N., Godec, J., Lau, L., Sivick, K.E., McLaughlin, L.M., Jones, M.B., et al. (2011) TLR signaling is required for *Salmonella typhimurium* virulence. *Cell* **144**: 675–688.

Bagg, A. and Neilands, J. (1987) Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. *Biochemistry* **26**: 5471–5477.

Barbet, A.F., Agnes, J.T., Moreland, A.L., Lundgren, A.M., Alleman, A.R., Noh, S.M., et al. (2005) Identification of functional promoters in the msp2 expression loci of *Anaplasma marginale* and *Anaplasma phagocytophilum*. *Gene* **353**: 89–97.

Baron, C. and Coombes, B. (2007) Targeting bacterial secretion systems: benefits of disarmament in the microcosm. *Infect Disord Drug Targets* **7**: 19–27.

Baumgartner, W.A., Dubytska, L., Rogge, M.L., Mottram, P.J., and Thune, R.L. (2014) Modulation of vacuolar pH is required for replication of *Edwardsiella ictaluri* in channel catfish macrophages. *Infect. Immun.* **82**: 2329–2336.

Blanvillain, S., MEYER, D., Boulanger, A., Lautier, M., Guynet, C., Denancé, N., et al. (2007) Plant carbohydrate scavenging through tonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS ONE* **2**: e224.

Breuer, W., Epsztejn, S., and Cabantchik, Z.I. (1995) Iron acquired from transferrin by K562 cells is delivered into a cytoplasmic pool of chelatable iron(II). *Journal of Biological Chemistry* **270**: 24209–24215.

Brickman, T.J., Cummings, C.A., Liew, S.-Y., Relman, D.A., and Armstrong, S.K. (2011) Transcriptional profiling of the iron starvation response in *Bordetella pertussis* provides new insights into siderophore utilization and virulence gene expression. *Journal of Bacteriology* **193**: 4798–4812.

Cascales, E. and Christie, P.J. (2003) The versatile bacterial type IV secretion systems. *Nat Rev Micro* **1**: 137–149.

Castaño-Cerezo, S., Bernal, V., Blanco-Catalá, J., Iborra, J.L., and Cánovas, M. (2011) cAMP-CRP co-ordinates the expression of the protein acetylation pathway with central metabolism in *Escherichia coli*. *Molecular Microbiology* **82**: 1110–1128.

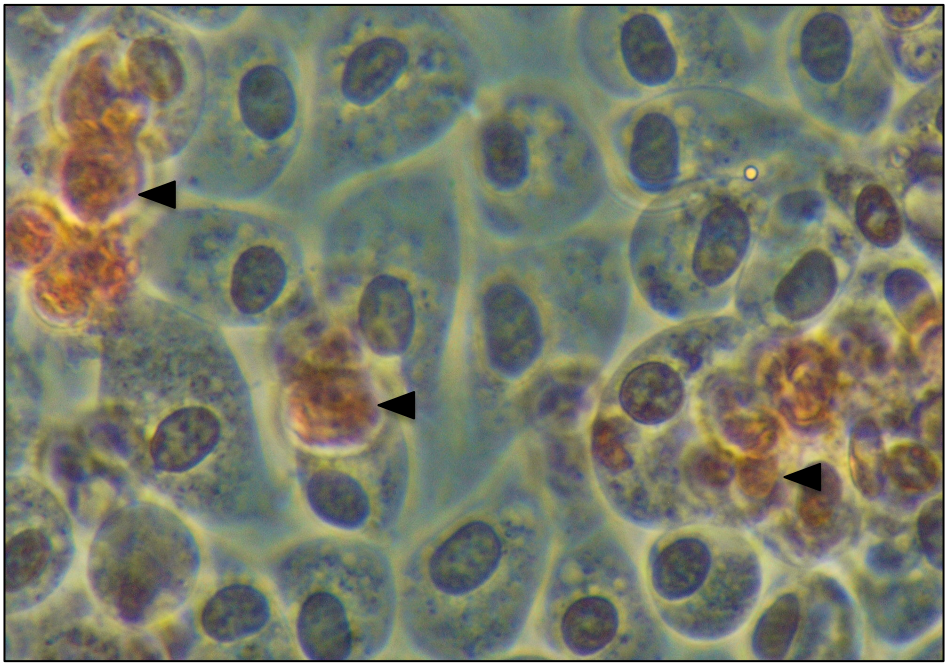
Chakraborty, S., Li, M., Chatterjee, C., Sivaraman, J., Leung, K.Y., and Mok, Y.-K. (2010) Temperature and Mg2+ sensing by a novel PhoP-PhoQ two-component system for regulation of virulence in *Edwardsiella tarda*. *J. Biol. Chem.* **285**:

38876–38888.
Cheng, Y., Liu, Y., Wu, B., Zhang, J.-Z., Gu, J., Liao, Y.-L., et al. (2014) Proteomic analysis of the *Ehrlichia chaffeensis* phagosome in cultured DH82 cells. *PLoS ONE* **9**: e88461.
Cheng, Z., Wang, X., and Rikihisa, Y. (2008) Regulation of type IV secretion apparatus genes during *Ehrlichia chaffeensis* intracellular development by a previously unidentified protein. *Journal of Bacteriology* **190**: 2096–2105.
Clague, M.J., Urbé, S., Aniento, F., and Gruenberg, J. (1994) Vacuolar ATPase activity is required for endosomal carrier vesicle formation*. *Journal of Biological Chemistry* **269**: 21–24.
Collins, N.E., Liebenberg, J., de Villiers, E.P., Brayton, K.A., Louw, E., Pretorius, A., et al. (2005) The genome of the heartwater agent Ehrlichia ruminantium contains multiple tandem repeats of actively variable copy number. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 838–843.
Coombes, B.K., Brown, N.F., Valdez, Y., Brumell, J.H., and Finlay, B.B. (2004) Expression and secretion of *Salmonella* pathogenicity island-2 virulence genes in response to acidification exhibit differential requirements of a functional type III secretion apparatus and SsaL. *Journal of Biological Chemistry* **279**: 49804–49815.
De Jong, M.F., Sun, Y.-H., Hartigh, den, A.B., van Dijl, J.M., and Tsolis, R.M. (2008) Identification of VceA and VceC, two members of the VjbR regulon that are translocated into macrophages by the *Brucella* type IV secretion system. *Molecular Microbiology* **70**: 1378–1396.
de Lorenzo, V., Wee, S., Herrero, M., and Neilands, J.B. (1987) Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (*fur*) repressor. *Journal of Bacteriology* **169**: 2624–2630.
Dumler, J.S., Barbet, A.F., Bekker, C.P., Dasch, G.A., Palmer, G.H., Ray, S.C., et al. (2001) Reorganization of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of Ehrlichia equi and "HGE agent" as subjective synonyms of *Ehrlichia phagocytophila*. *Int. J. Syst. Evol. Microbiol.* **51**: 2145–2165.
Dunning Hotopp, J.C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S.V., Eisen, J.A., et al. (2006) Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet* **2**: e21.
Escolar, L., Pérez-Martín, J., and de Lorenzo, V. (1998) Binding of the fur (ferric uptake regulator) repressor of *Escherichia coli* to arrays of the GATAAT sequence. *J. Mol. Biol.* **283**: 537–547.
Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. (2006) The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. *THE PLANT CELL ONLINE* **18**: 2680–2693.
Grifantini, R., Sebastian, S., Frigimelica, E., Draghi, M., Bartolini, E., Muzzi, A., et al. (2003) Identification of iron-activated and -repressed Fur-dependent genes by transcriptome analysis of Neisseria meningitidis group B. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 9542–9547.
Hyytiäinen, H., Sjöblom, S., Palomäki, T., Tuikkala, A., and Tapio Palva, E. (2003)

The PmrA-PmrB two-component system responding to acidic pH and iron controls virulence in the plant pathogen *Erwinia carotovora* ssp. *carotovora*. *Molecular Microbiology* **50**: 795–807.
IJdo, J.W., Carlson, A.C., and Kennedy, E.L. (2007) *Anaplasma phagocytophilum* AnkA is tyrosine-phosphorylated at EPIYA motifs and recruits SHP-1 during early infection. *Cell Microbiol* **9**: 1284–1296.
Jauron, S.D., Nelson, C.M., Fingerle, V., Ravyn, M.D., Goodman, J.L., Johnson, R.C., et al. (2001) Host cell-specific expression of a p44 epitope by the human granulocytic ehrlichiosis agent. *J. Infect. Dis.* **184**: 1445–1450.
Jones, C.M. and Niederweis, M. (2010) Role of porins in iron uptake by *Mycobacterium smegmatis*. *Journal of Bacteriology* **192**: 6411–6417.
Kumagai, Y., Huang, H., and Rikihisa, Y. (2008) Expression and porin activity of P28 and OMP-1F during intracellular *Ehrlichia chaffeensis* development. *Journal of Bacteriology* **190**: 3597–3605.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
Lee, J.-W. and Helmann, J.D. (2007) Functional specialization within the Fur family of metalloregulators. *Biometals* **20**: 485–499.
Li, Z. and Carlow, C.K.S. (2012) Characterization of transcription factors that regulate the type IV secretion system and riboflavin biosynthesis in *Wolbachia* of *Brugia malayi*. *PLoS ONE* **7**: e51597.
Liu, H., Bao, W., Lin, M., Niu, H., and Rikihisa, Y. (2012) *Ehrlichia* type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating host MnSOD. *Cell Microbiol* **14**: 1037–1050.
Llosa, M., Roy, C., and Dehio, C. (2009) Bacterial type IV secretion systems in human disease. *Molecular Microbiology* **73**: 141–151.
Lockwood, S., Voth, D.E., Brayton, K.A., Beare, P.A., Brown, W.C., Heinzen, R.A., and Broschat, S.L. (2011) Identification of *Anaplasma marginale* type IV secretion system effector proteins. *PLoS ONE* **6**: e27724.
Marcelino, I., Veríssimo, C., Sousa, M.F.Q., Carrondo, M.J.T., and Alves, P.M. (2005) Characterization of *Ehrlichia ruminantium* replication and release kinetics in endothelial cell cultures. *Veterinary Microbiology* **110**: 87–96.
Martín-Martín, A.I., Sancho, P., de Miguel, M.J., Fernández-Lago, L., and Vizcaíno, N. (2012) Quorum-sensing and BvrR/BvrS regulation, the type IV secretion system, cyclic glucans, and BacA in the virulence of *Brucella ovis*: similarities to and differences from smooth brucellae. *Infect. Immun.* **80**: 1783–1793.
Martínez-Núñez, C., Altamirano-Silva, P., Alvarado-Guillén, F., Moreno, E., Guzmán-Verri, C., and Chaves-Olarte, E. (2010) The two-component system BvrR/BvrS regulates the expression of the type IV secretion system VirB in *Brucella abortus*. *Journal of Bacteriology* **192**: 5603–5608.
Mekalanos, J.J. (1992) Environmental signals controlling expression of virulence determinants in bacteria. *Journal of Bacteriology* **174**: 1–7.
Mey, A.R., Wyckoff, E.E., Kanukurthy, V., Fisher, C.R., and Payne, S.M. (2005) Iron and fur regulation in *Vibrio cholerae* and the role of fur in virulence. *Infect. Immun.* **73**: 8167–8178.
Meyer, D.F., Noroy, C., Moumène, A., Raffaele, S., Albina, E., and Vachiery, N. (2013) Searching algorithm for type IV secretion system effectors 1.0: a tool for

predicting type IV effectors and exploring their genomic context. *Nucleic Acids Research* **41**: 9218–9229.
Nelson, C.M., Herron, M.J., Felsheim, R.F., Schloeder, B.R., Grindle, S.M., Chavez, A.O., et al. (2008) Whole genome transcription profiling of *Anaplasma phagocytophilum* in human and tick host cells by tiling array analysis. *bmc genomics* **9**: 364.
Niu, H. and Rikihisa, Y. (2013) Ats-1: a novel bacterial molecule that links autophagy to bacterial nutrition. *Autophagy* **9**: 787–788.
Niu, H., Kozjak-Pavlovic, V., Rudel, T., and Rikihisa, Y. (2010) *Anaplasma phagocytophilum* Ats-1 is imported into host cell mitochondria and interferes with apoptosis induction. *PLoS Pathog* **6**: e1000774.
Niu, H., Yamaguchi, M., and Rikihisa, Y. (2008) Subversion of cellular autophagy by *Anaplasma phagocytophilum*. *Cell Microbiol* **10**: 593–605.
O'Sullivan, D.J., Dowling, D.N., deLorenzo, V., and O'Gara, F. (1994) *Escherichia coli* ferric uptake regulator (Fur) can mediate regulation of a pseudomonad iron-regulated promoter. *FEMS Microbiology Letters* **117**: 327–332.
Oogai, Y., Matsuo, M., Hashimoto, M., Kato, F., Sugai, M., and Komatsuzawa, H. (2011) Expression of virulence factors by *Staphylococcus aureus* grown in serum. *Applied and Environmental Microbiology* **77**: 8097–8105.
Palyada, K., Threadgill, D., and Stintzi, A. (2004) Iron acquisition and regulation in *Campylobacter jejuni. Journal of Bacteriology* **186**: 4714–4729.
Porte, F., Liautard, J.P., and Köhler, S. (1999) Early acidification of phagosomes containing *Brucella suis* is essential for intracellular survival in murine macrophages. *Infect. Immun.* **67**: 4041–4047.
Portier, E., Zheng, H., Sahr, T., Burnside, D.M., Mallama, C., Buchrieser, C., et al. (2014) *IroT*/*mavN*, a new iron-regulated gene involved in *Legionella pneumophila* virulence against amoebae and macrophages. *Environ Microbiol*.
Postigo, M., Taoufik, A., Bell-Sakyi, L., de Vries, E., Morrison, W.I., and Jongejan, F. (2007) Differential transcription of the major antigenic protein 1 multigene family of *Ehrlichia ruminantium* in *Amblyomma variegatum* ticks. *Veterinary Microbiology* **122**: 298–305.
Pruneau, L., Emboulé, L., Gely, P., Marcelino, I., Mari, B., Pinarello, V., et al. (2012) Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development. *FEMS Immunol Med Microbiol* **64**: 66–73.
Rappl, C., Deiwick, J., and Hensel, M. (2003) Acidic pH is required for the functional assembly of the type III secretion system encoded by *Salmonella* pathogenicity island 2. *FEMS Microbiology Letters* **226**: 363–372.
Ratledge, C. and Dover, L.G. (2000) Iron metabolism in pathogenic bacteria. *Annu. Rev. Microbiol.* **54**: 881–941.
Ray, K., Marteyn, B., Sansonetti, P.J., and Tang, C.M. (2009) Life on the inside: the intracellular lifestyle of cytosolic bacteria. *Nature Publishing Group* **7**: 333–340.
Rikihisa, Y. (2010) *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*: subversive manipulators of host cells. *Nature Publishing Group* **8**: 328–339.
Rogge, M.L. and Thune, R.L. (2011) Regulation of the *Edwardsiella ictaluri* type III secretion system by pH and phosphate concentration through EsrA, EsrB, and EsrC. *Applied and Environmental Microbiology* **77**: 4293–4302.
Romeo, A.M., Christen, L., Niles, E.G., and Kosman, D.J. (2001) Intracellular chelation of iron by bipyridyl inhibits DNA virus replication: ribonucleotide

reductase maturation as a probe of intracellular iron pools. *Journal of Biological Chemistry* **276**: 24301–24308.
Sandrini, S., Masania, R., Zia, F., Haigh, R., and Freestone, P. (2013) Role of porin proteins in acquisition of transferrin iron by enteropathogens. *Microbiology (Reading, Engl.)* **159**: 2639–2650.
Shultis, D.D., Purdy, M.D., Banchs, C.N., and Wiener, M.C. (2006) Outer membrane active transport: structure of the BtuB:TonB complex. *Science* **312**: 1396–1399.
Skaar, E.P. (2010) The battle for iron between bacterial pathogens and their vertebrate hosts. *PLoS Pathog* **6**: e1000949.
Terradot, L. and Waksman, G. (2011) Architecture of the *Helicobacter pylori* Cag-type IV secretion system. *FEBS Journal* **278**: 1213–1222.
Trokter, M., Felisberto-Rodrigues, C., Christie, P.J., and Waksman, G. (2014) ScienceDirectRecent advances in the structural and molecular biology of type IV secretion systems. *Current Opinion in Structural Biology* **27**: 16–23.
Tseng, T.-T., Tyler, B.M., and Setubal, J.C. (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiology* **9 Suppl 1**: S2.
van Heerden, H., Collins, N.E., Brayton, K.A., Rademeyer, C., and Allsopp, B.A. (2004) Characterization of a major outer membrane protein multigene family in *Ehrlichia ruminantium*. *Gene* **330**: 159–168.
Wang, X., Cheng, Z., Zhang, C., Kikuchi, T., and Rikihisa, Y. (2007) *Anaplasma phagocytophilum* p44 mRNA expression is differentially regulated in mammalian and tick host cells: involvement of the DNA binding protein ApxR. *Journal of Bacteriology* **189**: 8651–8659.
Wang, X., Kikuchi, T., and Rikihisa, Y. (2007) Proteomic identification of a novel *Anaplasma phagocytophilum* DNA binding protein that regulates a putative transcription factor. *Journal of Bacteriology* **189**: 4880–4886.
Wang, Y., Chen, Z., Qiao, F., Zhong, Z., Xu, J., Wang, Z., et al. (2010) The type IV secretion system affects the expression of Omp25/Omp31 and the outer membrane properties of *Brucella melitensis*. *FEMS Microbiology Letters* **303**: 92–100.
Weber, G.G., Kortmann, J., Narberhaus, F., and Klose, K.E. (2014) RNA thermometer controls temperature-dependent virulence factor expression in *Vibrio cholerae. Proc. Natl. Acad. Sci. U.S.A.* **111**: 14241–14246.
Wu, C.-F., Lin, J.-S., Shaw, G.-C., and Lai, E.-M. (2012) Acid-induced type VI secretion system is regulated by ExoR-ChvG/ChvI signaling cascade in *Agrobacterium tumefaciens*. *PLoS Pathog* **8**: e1002938.
Yu, X.-J., McGourty, K., Liu, M., Unsworth, K.E., and Holden, D.W. (2010) pH sensing by intracellular *Salmonella* induces effector translocation. *Science* **328**: 1040–1043.
Zusman, T., Aloni, G., Halperin, E., Kotzer, H., Degtyar, E., Feldman, M., and Segal, G. (2007) The response regulator PmrA is a major regulator of the *icm/dot* type IV secretion system in *Legionella pneumophila* and *Coxiella burnetii*. *Molecular Microbiology* **63**: 1508–1523.
Zweygarth, E. and Josemans, A. (2001) Continuous in vitro propagation of *Cowdria ruminantium* (Welgevonden stock) in a canine macrophage-monocyte cell line. *The onderstepoort journal of veterinary research* **68**: 155–157.



Supplementary figure 1

A

AGTGTACCATTGATAATCGAGACAATTAAATTTATACTTAATTTTTTATAGTATCTTATGCAAGAAT

TCAAGCAATACAAATACTCTTACTTAAAAAAACGTATAAAAAAAATGTATTTTTTTATACATAAACTC

TCAAAAAGTGTAAATATATACTTCATCATCAATAAATTATAAAGTGATTTAATATAATATTTTTAAC

ATATGGATTAATTTATCAAGTATATATGAAAAAACAGTATTTGCTAACTCGTAACTGCATAAACAT

GTTTTATTAGTATAAAATATTTAAAATATTAATATTTATTTATAAGAAACAAGATAATAATATAAAA

ATTATAAGATTTTATGATCTGTAATATCATAGTTTTA**TATAAT**AACTAAG**TATTAT**TTTTTAAATTT

GA**TTG**

-35 **FUR box** **-10**

Codon initiation

B

Fur box *E. coli*	GATAATGA-TAATCATTATC
Fur box ErxR	TATAATAACTAAGTATTAT
Consensus	gATAATaA.TAAgcATTAT.

Supplementary figure 2

Partie 3

Caractérisation d'Erip1, un nouvel effecteur du SST4 d'*E. ruminantium*

1. Préambule

Lors d'une infection, les bactéries pathogènes prennent l'ascendant sur leur hôte, en contournant ses défenses. Pour y parvenir, elles injectent dans leur hôte des effecteurs de type IV (ET4s), qui leur permettent d'exploiter la machinerie cellulaire à leur profit et de déjouer les réponses immunitaires de l'hôte. Pour identifier les ET4s d'*E. ruminantium*, nous avons développé un logiciel S4TE (*Searching Algorithm for Type IV secretion system Effector proteins*). Ce logiciel prédit les ET4s candidats en combinant 13 caractéristiques présentes chez les ET4s connus. Cette stratégie a permis d'identifier 22 candidats potentiels chez *E. ruminantium*.

Des expériences de translocation par le SST4, en système hétérologue chez *L. pneumophila*, en utilisant le système CyaA ont permis d'identifier Erip1 (*Ehrlichia ruminantium* injected protein 1), le premier effecteur d'*E. ruminantium*. La partie centrale d'Erip1 contient un domaine NLS et sa structure tertiaire prédictive révèle des homologies avec des protéines nucléaires. Des expériences d'immunofluorescence pour déterminer la localisation subcellulaire d'Erip1 ont démontré qu'Erip1 était sécrété dans le cytoplasme et le noyau de la cellule hôte. Cette protéine possède également des domaines EPIYA-like en C-terminal et nous avons démontré qu'Erip1 était phosphorylé au niveau des tyrosines dans la souche Gardel virulente et atténuée. En revanche, cet effecteur semble clivé dans la souche atténuée et pourrait être en lien avec la perte de virulence.

Ainsi, sous réserve de l'identification de ses cibles nucléiques ou protéiques, cette protéine pourrait appartenir à la famille émergente des nucléomodulines, connues pour réguler la transcription ou la maturation d'ARN de la cellule hôte. Des expériences complémentaires seraient nécessaires pour identifier les interacteurs protéiques et les cibles d'Erip1 dans la cellule hôte, afin de déchiffrer le rôle d'Erip1 dans la pathogenèse due à *Ehrlichia*.

2. Manuscrit préliminaire: Erip1, a new substrate of the *E. ruminantium* type IV secretion system, is tyrosine phosphorylated and imported into host cell nucleus

Erip1, a new substrate of the *E. ruminantium* type IV secretion system, is tyrosine phosphorylated and imported into host cell nucleus

Amal Moumène[a,b,c], Adela Oliva Chavez[a,b], Silvina Gonzalez-Rizzo[d], Christophe Noroy[a,b], Christian Sheikboudou[a,b], Caylin G. Winchell[e], Thierry Lefrançois[b,f], Nathalie Vachiéry[a,b], Daniel E. Voth[e], and Damien F. Meyer[a,b*]

[a] CIRAD, UMR CMAEE, Site de Duclos, Prise d'eau, F-97170 Petit-Bourg, Guadeloupe, France

[b] INRA, UMR1309 CMAEE, F-34398 Montpellier, France

[c] Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

[d] UMR 7138, équipe Biologie de la Mangrove, Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

[e] Department of Microbiology and Immunology, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

[f] CIRAD, UMR CMAEE, Montpellier, France

[*]Corresponding author: damien.meyer@cirad.fr

Erip1, a new substrate of the *E. ruminantium* type IV secretion system, is tyrosine phosphorylated and imported into host cell nucleus

Amal Moumène[a,b,c], Adela Oliva Chavez[a,b], Silvina Gonzalez-Rizzo[d], Christophe Noroy[a,b], Christian Sheikboudou[a,b], Caylin G. Winchell[e], Thierry Lefrançois[b,f], Nathalie Vachiéry[a,b], Daniel E. Voth[e], and Damien F. Meyer[a,b*]

[a] CIRAD, UMR CMAEE, Site de Duclos, Prise d'eau, F-97170 Petit-Bourg, Guadeloupe, France

[b] INRA, UMR1309 CMAEE, F-34398 Montpellier, France

[c] Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

[d] UMR 7138, équipe Biologie de la Mangrove, Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France

[e] Department of Microbiology and Immunology, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

[f] CIRAD, UMR CMAEE, Montpellier, France

[*]Corresponding author: damien.meyer@cirad.fr

Abstract

Bacterial pathogens have evolved numerous strategies to corrupt, hijack, or mimic host cellular processes to survive and proliferate. *Ehrlichia ruminantium* is an obligate intracellular bacterium that causes heartwater, a fatal and economically important disease of wild and domestic ruminants. This bacterium uses a type IV secretion system (T4SS) predicted to play an important role in invasion and pathogenesis. T4SSs are specialized ATP-dependent protein complexes that deliver type IV effector (T4E) proteins into eukaryotic cells to subvert host cell processes during infection. To identify *E. ruminantium* T4Es, we developed a software named S4TE (*Searching Algorithm for Type IV secretion system Effector proteins*). This tool predicts and ranks T4E candidates by using a combination of 13 sequence characteristics, including homology to known effectors. This strategy identified 22 potential *Ehrlichia ruminantium* T4Es. Using the *L. pneumophila* T4SS model with an adenylate cyclase (CyaA) enzymatic reporter, we validated the T4SS-dependent translocation of the first *E. ruminantium* effector Erip1 (*Ehrlichia ruminantium injected protein 1*). Erip1, the biggest known T4E to date (460 kDa), is specific to *Ehrlichia* and show no other homologies to any bacterial protein in the database. The central part of Erip1 contains a nuclear localization signal and the predicted tertiary structure revealed potential homologies to nuclear proteins. Analysis of subcellular localization of Erip1 showed that it is secreted in the cytoplasm with a peak of expression at early stages of bacterial development, before exponential growth. Erip1 is also tyrosine-phosporylated inside the cytoplasm of infected cells and translocated in the nucleus at 72 hpi. This protein may be part of the emerging family of the nucleomodulins, proteins injected into the nucleus of host cells that regulate gene transcription or RNA maturation. Identification of protein interactors and targets of Erip1 will provide valuable information about its function and molecular mechanisms underlying *Ehrlichia* pathogenesis and host cell response.

Introduction

Ehrlichia ruminantium, the causative agent of heartwater, is an obligate intracellular Gram-negative alpha-proteobacterium, belonging to the family *Anaplasmataceae* order *Rickettsiales* that develops within membrane-bound vacuoles (Dunning Hotopp et al., 2006; Allsopp, 2010). Heartwater is distributed in nearly all countries of sub-Saharan Africa from where it has been imported into some islands of the Caribbean along with its vector, the tick *Amblyomma variegatum* (Barré et al., 1987). The risk of spread into the American mainland is increased as infected ticks have been shown to feed on migratory birds (Vachiery et al., 2013). Within the mammalian host and tick vector, *E. ruminantium* organisms infect different cellular environments. In the mammalian host, *E. ruminantium* infects neutrophils and reticulo-endothelial cells, whereas in the tick-vector, it develops mostly in the midgut and salivary glands (Prozesky and Plessis, 1987). *E. ruminantium* has two forms during its developmental cycle. The replicative form resides within the vacuole where it divides by binary fission and the extracellular infectious form known as elementary body (Jongejan et al., 1991). In spite of the growing knowledge about the factors that are essential for *Ehrlichia* spp. intracellular replication, little is known about what molecular determinants allow *E. ruminantium* invasion and persistence in eukaryotic cells.

Pathogenic bacteria have evolved a multitude of virulence factors in order to evade the host immune responses and to reach a replicative niche. The type IV secretion system (T4SS) is involved in the injection of virulence factors into target cells by several mammalian pathogens (Cascales and Christie, 2003). Rapid progress has been made towards identifying the proteins that form part of the different T4SSs, the translocated effectors and how the effectors subvert eukaryotic cellular processes during infection (Voth et al., 2012). In *Anaplasmataceae*, several effectors have been discovered. In *Anaplasma phagocytophilum*, only two T4SS substrates have been identified and partially characterized. One is the ankyrin repeat domain-containing protein (AnkA), which has been shown to translocate to the host nucleus and interact with DNA affecting host innate immunity (IJdo et al., 2007). The second is known as *Anaplasma* translocated substrate 1 (Ats-1), whose C-terminal portion is imported into the mitochondria where it is proposed to interfere with the induction of apoptosis

(Niu et al., 2010). The full length Ats-1 also hijacks autophagy pathway to acquire host nutrients (Niu et al., 2010). There is also evidence that the *E. chaffeensis* T4SS is used to secrete bacterial effector proteins into the host cytoplasm that then translocated into the mitochondria (Liu et al., 2012). T4SS structural proteins are definitively expressed in *A. marginale* within host erythrocytes (Sutten et al., 2010) and recent studies have identified several T4SS effector molecules in this bacterium, including AnkA (Lockwood et al., 2011). Similarly, it is suspected that effector proteins play an important role in invasion and pathogenesis in *E. ruminantium* (Collins et al., 2005; D. F. Meyer et al., 2013).

Protein phosphorylation at tyrosine residues is important in eukaryotic cells signal transduction (Pawson and Scott, 2005). Thus, several intracellular organisms secrete bacterial virulence factors into the host cell that then undergo tyrosine phosphorylation and interfere with signal transduction pathways. A well-studied example of this is the *H. pylori* CagA T4SS effector, which is tyrosine phosphorylated at EPIYA motifs within the protein (Hatakeyama, 2004). Once in the host cell, CagA interacts with host cell proteins affecting their tyrosine phosphorylation, which leads to deficiencies in actin rearrangements, triggering of inflammatory responses and the development of gastric carcinogenesis (Hatakeyama, 2004; Backert and T. F. Meyer, 2006). Several other "bacterial EPIYA effectors" have been identified, such as *Anaplasma phagocytophilum* AnkA (IJdo et al., 2007), enteropathogenic *Escherichia coli* Tir (Campellone et al., 2002), *Citrobacter rodentium* Tir (W. Deng et al., 2003), *Chlamydia trachomatis* Tarp (Clifton et al., 2004; Mehlitz et al., 2008), *Haemophilus ducreyi* LspA (K. Deng et al., 2008), and *Bartonella henselae* BepD, BepE, and BepF (Selbach et al., 2009). In this study, we demonstrated that Erip1, named *Ehrlichia ruminantium* injected protein 1 is an *E. ruminantium* T4SS effector that is injected into the host cell's cytoplasm and then transported into the nucleus. Bioinformatic approaches determined that the protein shares structural similarities with protein belonging to the family called 'nucleomodulins'. This family of proteins interfers with the host DNA or several host nuclear proteins affecting transcription, chromatin remodeling, and DNA replication (Bierne and Cossart, 2012). Erip1 is a protein of 460 kDa that contains EPIYA-like motifs in the C-terminal region of the protein and is tyrosine phosphorylated. This and the nuclear localization of the protein suggest that

Erip1 may interfere with host signaling.

Materials and methods

E. ruminantium cultivation and expression analysis

E. ruminantium Gardel virulent and attenuated strains were routinely propagated in bovine aortic endothelial (BAE) cells as previously described (Marcelino et al., 2005). *E. ruminantium* samples were collected from two 25 cm^2 flasks at each time point at early (24 hpi = T1 and 48 hpi), intermediate (72 hpi = T2 and 96 hpi), and late stages of development (120 hpi = T3). Each flask was divided at each time point post-infection and 1/10 (600 µl) of sample was centrifuged at 20,000 X g for 5 min for DNA extraction, using QIAamp DNA Mini Kit (QIAGEN, France). The number of bacteria in each flask at each time point was quantified by q-PCR (Emboulé et al., 2009). The remaining 9/10 (5400 µl) from each flask were retained for either RNA or protein extraction. Total RNA extraction procedure was performed as described by (Emboulé et al., 2009).

CyaA protein translocation assay

Several predicted effectors were identified using S4TE program (D. F. Meyer et al., 2013). The full length of each putative effectors or a 300 bp portion of the C-terminal from each gene were amplified from genomic DNA by PCR using AccuPrime™ Pfx DNA polymerase (Invitrogen) and gene-specific primers containing BamHI/SalI specific sites (Table 1). The resulting PCR products were cloned into the plasmid pJB2581 using the Infusion Kit (Clontech). All plasmid inserts were sequenced to verify individual clones. *L. pneumophila* CyaA translocation assay was used to test the secretion of each effector as described by Lockwood *et al.* (Lockwood et al., 2011). *L. pneumophila* expressing CyaA alone was used as negative control and results of translocation were expressed as fold change over cAMP levels in this control. Experiments were performed in 6 replicates. Error bars represent the standard deviation from the mean. P-values (<0.01) were calculated using a Student's *t*-test. Statistical analyses were performed using GraphPad Prism 5 software (GraphPad Software, San Diego, CA, USA).

Structure Prediction

A secondary structure of Erip1 protein was predicted using S4TE (*Searching*

algorithm for type-IV secretion system effectors) (D. F. Meyer et al., 2013). We used I-TASSER tool to produce a high quality 3D model of Erip1, as described by Zhang (Zhang, 2008). The putative function of Erip1 and the identification of putative ligand binding sites were predicted from the previously generated 3D model of Erip1, using COFACTOR (A. Roy et al., 2012).

Western blot conditions for Erip1

We used Western blots to analyze the protein expression and the phosphorylation of Erip1. Protein samples were centrifuged at 4500 tr/min for 10 min. The supernatant was collected and centrifuged at 20,000 X g for 10 min. The pellet was re-suspended in PBS containing an anti-protease cocktail (50X) (Roche, Germany). The sample was centrifuged again at 20,000 X g for 5 min, re-suspended in SDS 2X (200 mM Tris, 20% Glycerol, 1.6% SDS, 0.02% Bromophenol blue, and 4% beta-mercapthoethanol), and boiled for 5 min. We used different volumes depending of the pellet size of the sample. The protein concentration from all samples was measured using the BCA kit 23235 (Pierce, USA). Protein samples (10 µl) were loaded into 4 – 12% NuPAGE Bis Tris Pre-cast gel ladder (Invitrogen, UK). The HiMark pre-stained ladder (Invitrogen, UK) (7 µl) was used as size reference. The gel was run for 3 h at 70 – 80 V to let the high molecular weight proteins migrate. Proteins were transferred overnight into Amersham Hybond-P membrane (GE healthcare limited, UK) in a TE 77 ECL Semidry transfer unit (Amersham biosciences, UK) at 20 V making sure that the filter papers and the unit were wet. The membrane was washed in PBS three times for 5 min and blocked in 5% milk in PBS for 1 h. After three 10 min washes in PBS, the proteins were detected using rabbit polyclonal clonal antibodies against peptide 1410706 (peptide sequence: NTRVSKTSSRRYVSN) named anti-Erip1 monoclonal herein, at 1:20,000 dilution (Eurogentec, France) for 3 h at RT. Membrane was washed three times for 10 min and labeled using Goat anti-rabbit Alkaline phosphatase (AP) (Invitrogen, UK) at a 1:500 dilution for 1 h and then detected with with NBT/BCIP (Roche, Germany).

For the detection of tyrosine phosphorylation of Erip1, proteins were run as described above and transferred to a membrane. The membrane was incubated in 3% Bovine Serum Albumin (BSA) in PBS with 1:1000 dilution of mouse monoclonal 4G10 Platinum mouse Anti-Phosphotyrosine (Merk Millipore) for 2 h at RT. Membrane was

washed three times for 10 min and proteins were detected with Rabbit anti-mouse Alkaline phosphatase (AP) labeled antibodies (Sigma-Aldrich, Missouri, USA). Bands were stained with NBT/BCIP as described above.

Confocal microscopy

BAE cells were grown in eight-well chamber slides as previously described (Marcelino et al., 2005). After three days, the cells were infected with *E. ruminantium*. *Ehrlichia*-infected cells were fixed in 2% paraformaldehyde at RT for 2 h. After three washes, the cells were permeabilized with 0.1% Triton X-100 for 15 min. The cells were blocked with 3% BSA at RT for 1 h. Cells were then incubated with rabbit anti-Erip1 monoclonal antibodies (1:20,000) for 2 h and Alexa Fluor 647-conjugated goat anti-rabbit IgG (1:500, ab150079, Abcam, UK) as secondary antibodies for 1 h. The samples were washed three times with PBS and then the cells were stained with SYBR Green I (1:10,000, Sigma-Aldrich, France) for 10 min before examination by confocal microscopy. The slides were examined with a Leica DM2500 laser scanning confocal microscope (Leica, Germany). The images were analyzed using ImageJ (National Institute of Health, USA).

Results

Choice and characteristics of six putative *E. ruminantium* T4S effectors

To identify potential T4SS effector proteins in *E. ruminantium*, we used the list of putative effectors previously predicted by the computational tool, S4TE (D. F. Meyer et al., 2013). The tool compares the protein sequence of each candidate to other known effectors and looks for several characteristics within the sequences, such as eukaryotic domains, subcellular localization signals or secretion signals, etc. After the prediction, each effector is ranked by score depending on the numbers of features within the protein sequence (D. F. Meyer et al., 2013). S4TE provided a list of 22 putative effectors for *E. ruminantium* (D. F. Meyer et al., 2013). We selected six proteins, ERGA_CDS_00570, ERGA_CDS_02150, ERGA_CDS_03640, ERGA_CDS_03830, ERGA_CDS_04230 and ERGA_CDS_06470, based on the ranks obtained with S4TE prediction and the total hydropathy score. The majority of these proteins have total hydropathy score less than -200 (going from -244 to -955). This criterion was recently used to screen the *A. marginale* proteome and resulted in

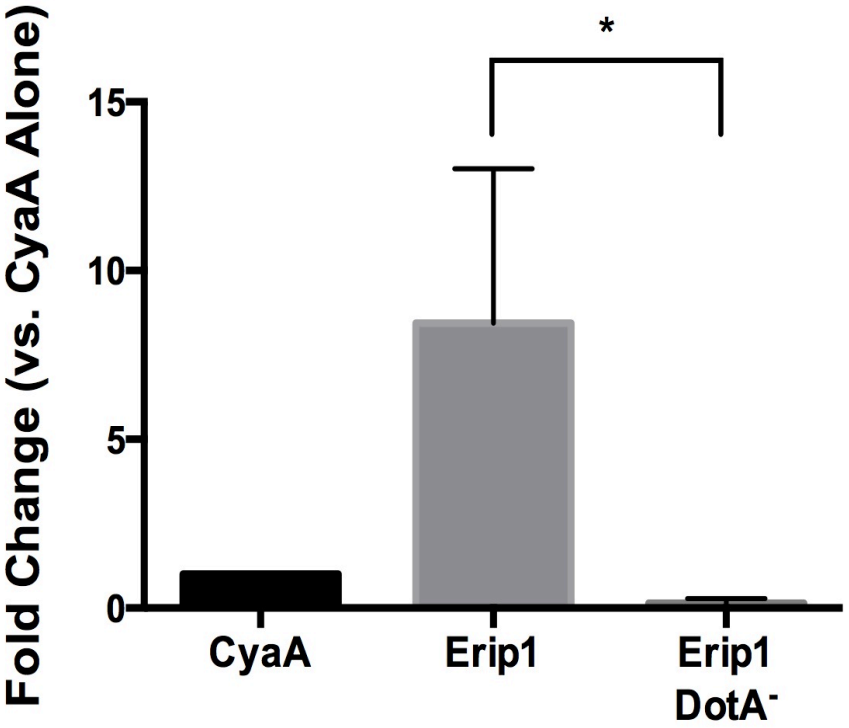
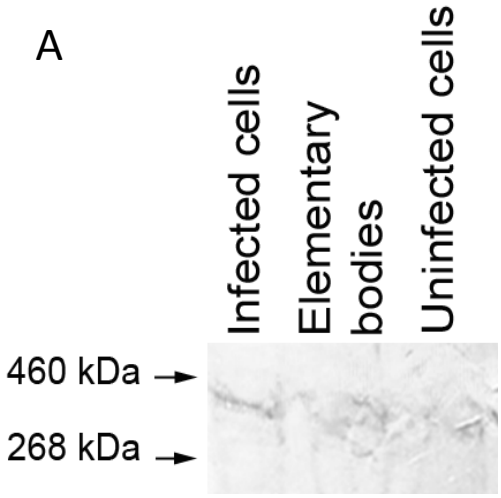


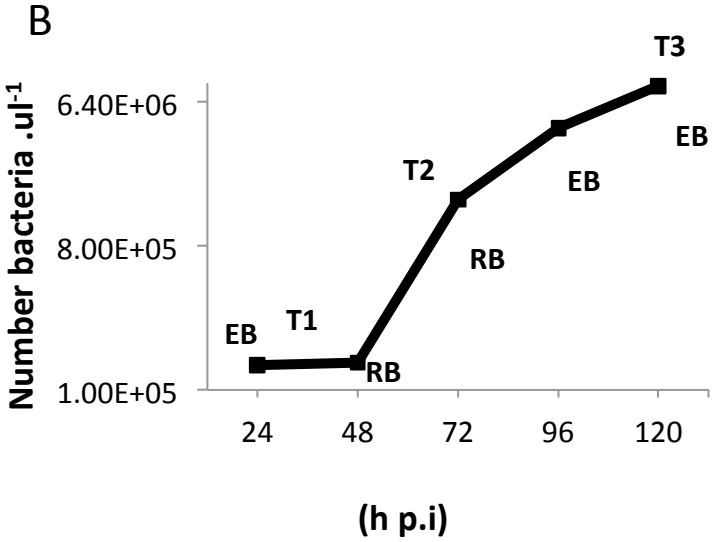
Figure 1

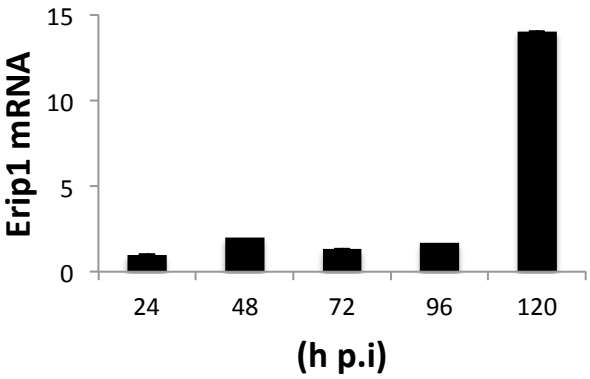
the identification of 33 putative effectors (Lockwood et al., 2011). In addition, the C-terminal of T4Es is frequently rich in alkaline amino acids and posses a positive charge (Vergunst et al., 2005; Rikihisa and Lin, 2010; Niu et al., 2010). We looked for the net positive charge of ERGA_CDS_00570 and ERGA_CDS_02150, which should be at least 2 as it has been shown in other *alpha-proteobacteria*. Another important factor is the presence of subcellular localization signals. Nuclear localization (NLS) and mitochondrial localization (MLS) signals were found in ERGA_CDS_00570 and ERGA_CDS_06470 respectively. Additionally, ERGA_CDS_03830 contains ANK domains and is predicted to be a homolog to *A. marginale* AnkA, a protein that is translocated to the nucleus in a T4SS-dependent manner. ANK was used for positive control since proteins bearing ankyrin repeats have a high probability of being secreted by the T4SS (Pan et al., 2008; Voth et al., 2009). ERGA_CDS_06470 is predicted to be an exodeoxyribonuclease V beta chain *recB*, whereas ERGA_CDS_00570 and ERGA_CDS_02150 encode hypothetical proteins.

Secretion of *E. ruminantium* ERGA_CDS_00570

The *Bordetella pertussis* CyaA reporter system has been used to study the translocation of many bacterial effectors into eukaryotic cells (Lockwood et al., 2011). We used this system to determine the translocation of the candidate effectors in *E. ruminantium*. The full coding sequence or portion of the C-terminal containing 100 amino acids for each of the identified effectors were fused to *B. pertussis* adenylate cyclase CyaA gene. For ERGA_CDS_00570, only the C-terminal portion of was cloned into the CyaA construct as the full protein is too large (3448 amino acids) and it is the portion recognized by the Dot/Icm system of *L. pneumophila* (Lockwood et al., 2011). We checked the correct expression of the proteins in *L. pneumophila* by SDS-PAGE and all proteins were expressed at the correct size (fig. S1). The constructs were cloned into *L. pneumophila* and the bacteria were infected into THP-1 cells. The level of cAMP was measured to determine the translocation of the effectors. cAMP is only produced after the translocation of the fused CyaA into the cytoplasm of the host cell where it interacts with calmodulin. Only the CyaA-ERGA_CDS_00570 fusion generated significant cAMP production (10 fold higher) when compared to controls, the CyaA alone and the DotA$^-$ mutant (figure 1). The


A
Infected cells
Elementary bodies
Uninfected cells
460 kDa
268 kDa


B
Number bacteria .ul-1
6.40E+06
8.00E+05
1.00E+05
T3
EB
T2
EB
RB
EB
T1
RB
24
48
72
96
120
(h p.i)


C
Erip1 mRNA
15
10
5
0
24
48
72
96
120
(h p.i)

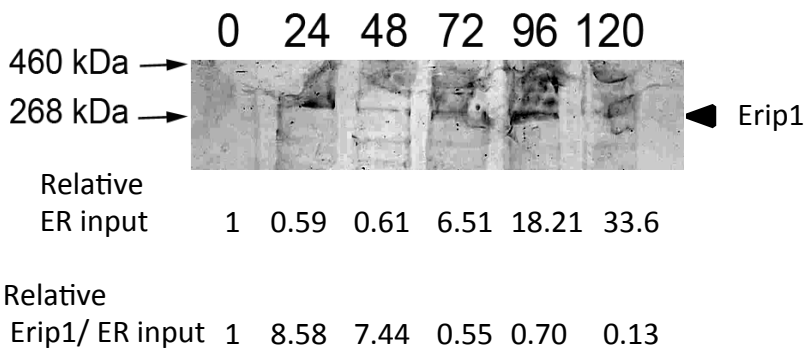

D
0 24 48 72 96 120
460 kDa
268 kDa
Erip1
Relative ER input 1 0.59 0.61 6.51 18.21 33.6
Relative Erip1/ ER input 1 8.58 7.44 0.55 0.70 0.13

Figure 2

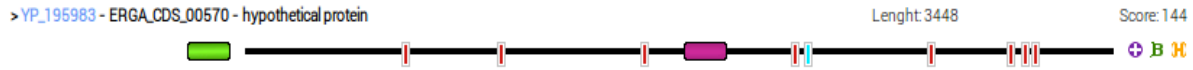
ERGA_CDS_00570 ORF encodes a 460 kDa protein that we have named *E. ruminantium* injected protein 1 (Erip1). The five other remaining proteins didn't show significant production of cAMP (fig. S2).

Erip1 is up-regulated early during exponential growth

To confirm expression of Erip1 during infection of BAE cells, we developed an Erip1-specific antibody against a 15 amino acid peptide (NTRVSKTSSRRYVSN). When tested by Western blot on lysates from *E. ruminantium*, the anti-Erip1 antibody revealed a major band at the expected size of 460 kDa (figure 2A) and a faint band with lower molecular weight (data not shown). No bands were detected in uninfected BAE cells, confirming the specificity of the Erip1 antibody (figure 2A). As mentioned above, *E. ruminantium* has a complex lifecycle and undergoes a biphasic development (Jongejan et al., 1991). We traced the different development stages with real-time PCR and determined that the lag phase occurred from 0 to 24 hpi, the pre-exponential growth phase from 24 to 48 hpi, the exponential growth phase from 48 to 96 hpi, and the stationary phase from 96 to 120 hpi. The phases of pre-exponential and exponential growth correspond to the time when the bacteria are in the reticulated form, whereas the lag and stationary phase correspond to the elementary body form. Under our culture conditions, the number of bacteria increased over 600-fold at 120 hpi compared with 0 hpi (figure 2B). We looked at the mRNA expression of Erip1 using qRT-PCR. The amount of *erip1* transcript, normalized by the bacteria number, was highest at 120 hpi with a small secondary peak at 48 hpi (figure 2C). To determine if Erip1 protein levels showed a similar trend as that detected by qRT-PCR, we looked at Erip1 concentrations at different time points by Western blotting. In all time points, the antibody recognized the protein at 460 kDa corresponding to the size of predictive Erip1. Protein band density for each sample was measured by densitometry using ImageJ and the ratio of protein was normalized to the bacterial number. Erip1 protein expression relative to bacterial DNA peaked at 24 and 48 hpi, corresponding to the exponential growth stage (figure 2D).

Analysis of T4S effector features for Erip1

The analysis using S4TE software showed that Erip1 harbors several characteristic features of proteins that may be secreted by the T4SS. The protein showed nuclear



MLNQKLDMAYIERLISQQPQLKVVNAETVSLSYSTVWDYIFNIELVQSIGYNAYVKNIRAYHPDYNVFHGVILCNPQVFLEEI
KIPRYGGILSKIQKDYPDCESLELYILNGYVLSVLVNDPLFEQFKSKITSIKMGDIDLFKQVYNDLKQNPNGHSEEDIDFLVR
GVMTDTHIRKYISLFPSLKELDCVISEPKCCETLVDVLSDVPILSQLSLRCARTSTSVEHVSKKDIKVRQQVEKLLLINCPME
ILLYFHGIRFLFITAPQNVSDSKVILLEKYCDNPSLLEVLLVSTNTYHQMQDGSIDFLSIDYDVSVFNNLKLLIINPCQDIRA
ILSVDGLFDSKAGLDFCNELVTPRKYSTLFNNGKPYPKSIEQVMILSSSLHQFSLLTDSEKLLISGLTSLFVSTTLIRPCLDI
NTLVPYLMYAVDSIQDGIINPDLNEQENVLLISLINQVKRKVDDLLLTLLEDSRLCDDLASYLKRMEFDDMVLTNASSVYSLS
YEPSVLFAINIDIPRISIEFLKKISNINLAILCKISLSDDVNSSTYLALTTFSAIRSIIRHTIGMSKFYVTTMKAWDYFINSM
AIGDRDLDLARVISAVFVQPYDTESVRDVCRQTVVGIKEHSEITYKRLEAENVNIFPQEQQVVEFIAGCVVFHSMYLIITNSA
ERLETQSCSPSLFIEILRDNVVNVCNNSLFDILKFDDCLLQIHYMIKQSFNHRPALTSDEVYTFASQCMINEEMLLQDVKNQG
FNKIDIRILKSSLFLIKVLDYIGYCPDVLFDQLQSYQLQTPMMILNVIFKTVSDLIKKRCESMQANEYTVFTLCYIINKAVSN
GSVSVRQLHKLLFSVIIDAVPQIACDVKYVKTNIHVLIGNMIKLCYFDYSYLEDKVYQKALLHIACFIVNFSITAARYEAQSD
CFIDEAVFLGFPDYVIDEIRLLGGIPHRSILQVLVRLLLKHYVWRVMEGVSGLNKFDYSWVCLQNMMQYKSPYDIQHISFFPT
IEDFEFLYSGILCNTYLPSKIEFEFAIRYLYPVISFGTKYICESISVIMADLLLKRMVCSYANAYTIEEKKLMRKLLLRALYK
FQEYEFPNFIEYMIKVSNVFDYEIEQFIYEEPLLHDLNFVWILKNVKPSEIINVIDKNIAEYREASDHMKLYDMIEVYIRQSY
KVPNLPSMFTDMQFIKCCGKILNVISTLHSRLGYLSNNHFIEMMQVVMTTYKISMQKPEYLGNCILNVQYCRDKVGGVLEECR
DGLDKVVSCLSNDNKVIDIIAIKDDLLSKCQDVEFIKTLNYRTVALYLKPYMVPPVQYSSDIVSHVIFCVVGEIIEQHNKREI
FFSQLTSALMLKKKLNTMQSVSEVVLEVFIKSLESNEEVTQGKEVSSTKQSKQKKISFDAYINDLLNKYFCKWWYTILDNMLA
GESLNHSIDLYTLKLFISEENLPVKVELALQNKIENEETNQYLVLINKFKSICFYQGNATNYFSDCSFQKRVLHNIITKIRYP
ATSSDKISVLSNRVIEQGSILLPEGLDHGKLQLAGKLYCATGTNDCHIKYIHLLQVRLFSEFFSHNHNALDIDVLDEKAYEVL
SISNASREYDCNNVVGFFVNMPSCLQDGQLDGLDPCSPKVINVLGLSLVWDTVSMLLHDLVFGNYLNDISRILCYVEDRFRIE
NQIDSSVLLLIKFHYKAILFKYEKGNYKDNHPIIFLHKFLQSFNYDLLEYYAPGCRVSLQPVVKKQSEVVTTEASSEVVKKAN
KKKSHSKKGKDKEVKQAVPETQDEKLVSGEDDKKSQAVVEKQLCVVHETVGADGVHLNPKDQDECYEGVVGDSVKLGDTTARL
LITTVPNNECLEQQDKPKVKKSVSHSKKKSKKSTAGCSTKTITSQQDKVNKSAVVEDVASAMLLTGDDTKVSESSQEQGVDEK
GKGKLCVVDTSGLVVSGQQGKDEESSIINKAVKEVVPDTKDEKLVSSGKKSKDVSTEGQLCVVPETADVDHVPLNPKGQDEYY
EGVVTDSVKLDDFTAGQLLITTVLNDEYYEAVECDGVQVNDSTVSRSLVDPVSGDQYVKQGARPKVKGSLSHNKRKKSKKSTE
ICATSTVKSQKDDLDKSAVVTTEKVDNTALLASVDVGVSELEQEKFTCEKDEYKFQIMEDTNAPAVLDKQQDSEEKRTIDNTS
KVLLESSHTTEQDYVISEESITQLCTSDDTLNLVDGSSQSLEACKDEQTVDDVAGNILQGKEKKQSQLTKSQRQRANKKARKI
KEQKEMVDVQKDKPELVECSELKTEEEDTINLLSCFDARVEPVIAYHPMSLEYVLNKYRVTNDSMSKTRFSHLLYYLRLTRSL
FMPVKGQETSVLVVTESAEKVLHDFFPEKGDYDFFLSMIKLSILTNFLMIIRPYSKFEAMTALFKHCYKIFTMDLDAFLCDLI
QEMYCISFQDVSLEKLRNHVLKVFEKYESQFCWNRSSAVTKYPLYFLSRVLSICPLISQAQHTSVSELQHVICNAGIICIVTH
IGHMMQQIGHMVEHSDGILYNNLKLLRRSGRSSDTVFDFIQQCYYFRNASCHLYDGKGASSNLNFFRTTSIIVDQFQVFIQSN
NGSFIDVIYDFFRQYDDVISKKNEICSKQQSSISVIEVDATKLAPRVINHEESTEGLFSELSNSGSSLQMQECFPEISKSLSL
GSNDVMRDTSEEDVSSKGTEHQGDALALMETRGSTSTDQVIDSVAVQEQEYQYSYLLQNLALQEKREASSSYLARSASGVGDS
SAVQYESVSGMFSQFSEDDLQNMRVVNSTSQSKIPATVGAYSCCDETVPINPSSELYDNMCMGSQRYSNDSMDSTDRWLAGQG
NNYSMGANSDYRINPESQKSTSSIMSASTSSSGFSLQQGSANEEKSGVESFGQDLQLHHMDSQGECGLLDSAWYGGGTDYLAS
TNSGVFNFQESSADEGSSEMLMKHYDQSLEHRQHSEYRLFDSTHQDSVDDGNIFSLRRNNFDEISNRVVGGRFENLPDTRMDF
SYEERLRNLSLHQGDGYNDVALASNDVLGLRQEGPYRMDRPVVGGRFENLPDTRMDFSYEERLRNLSLHQGDGYNDVALASNN
VRSLPQEGPYRMDRPVAGAFARNLSGTQLDQSSDRDVALASSNVRSLPQEGPYRIDRPVAGAFARNLSGTQLDQSSDRDVALA
SNYVLSPSMVSDLITQVFARNLQRNRVDHSSYEYRVHDPLLHHSSSRNIALASNDVPSLQQNSPDRVNNPDLPSSYMHSSREY
RVHSPLLYQSGDSNVASVSDNMSSLQQNSPDRVGNPDLPSSYMHSSREYRVHNPLLYQSDDSNVASVSDNMSSLQQNSPGRVD
SLFVEGCTPSLEHGLVGGHDELHNPLLCEKIGDNSFSSSVRSRGPRFYKEKSHQSPIGPRFCKARSHLTTHLASTELSDVDYQ
SGQHPVTGNLIQNNPNNPSNRRSRIPLSLQNTRVSKTSSRRYVSN

Promoter:	Name	Start	Sequence	End	Score	Strand
	-> pmrA	-71	GTAAATAAATAATTAATATT	-51	0.75501	+

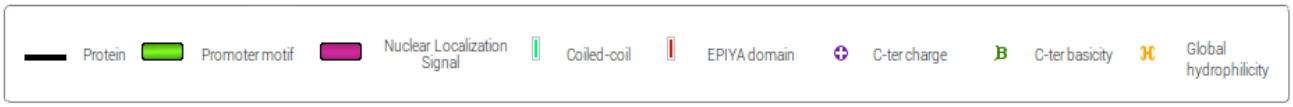
NLS:	Type	Start	Seq	End
	Bipartite	1740	KKANKKKSHSKKGK	1754

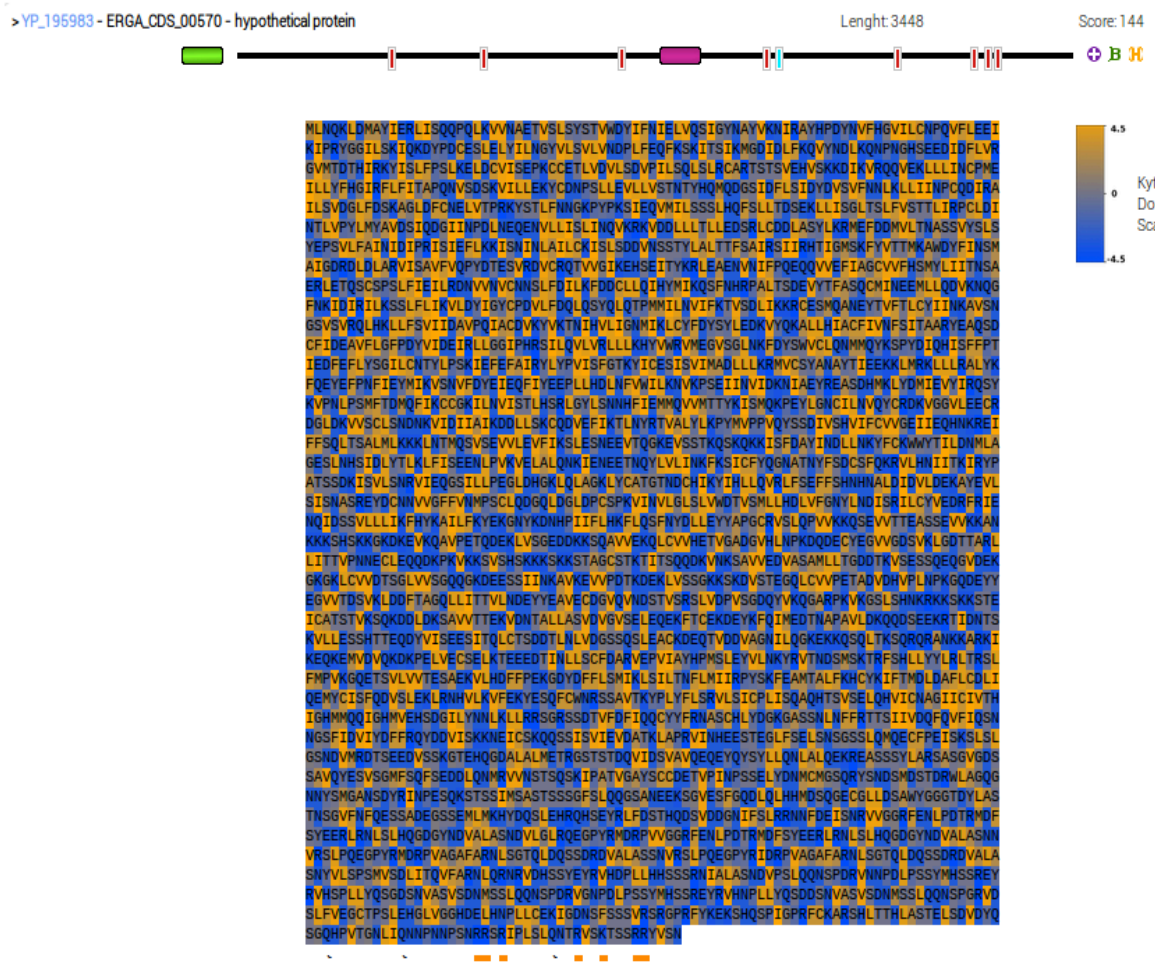
EPIYA:	Type	Start	Seq	End
	Hypothetical	623	EITYK	628
	Hypothetical	1001	EFLYS	1006
	Hypothetical	1571	EKAYE	1576
	Hypothetical	2169	EQDYV	2174
	Hypothetical	2704	EQEYQ	2709
	Hypothetical	3021	EGPYR	3026
	Hypothetical	3078	EGPYR	3083
	Hypothetical	3121	EGPYR	3126

Coiled-coil:	Score	Start	Seq	End	Proba
	1.494	2218	GKEKKQSQLTKSQRQRANKKARKIKEQKEMVDVQKDK	2254	0.943

Charges: 6 Basicity: 7 Cter Hydrophobicity: 0 Global Hydrophilicity: -955.6

Figure 3A





Promoter:	Name	Start	Sequence	End	Score	Strand
	-> pmrA	-71	GTAAATAAATAATTAATATT	-51	0.75501	+

NLS:	Type	Start	Seq	End
	Bipartite	1740	KKANKKKSHSKKGK	1754

EPIYA:	Type	Start	Seq	End
	Hypothetical	623	EITYK	628
	Hypothetical	1001	EFLYS	1006
	Hypothetical	1571	EKAYE	1576
	Hypothetical	2169	EQDYV	2174
	Hypothetical	2704	EQEYQ	2709
	Hypothetical	3021	EGPYR	3026
	Hypothetical	3078	EGPYR	3083
	Hypothetical	3121	EGPYR	3126

Coiled-coil:	Score	Start	Seq	End	Proba
	1.494	2218	GKEKKQSQLTKSQRQRANKKARKIKEQKEMVDVQKDK	2254	0.943

Charges: 6 Basicity: 7 Cter Hydrophobicity: 0 Global Hydrophilicity: -955.6

Figure 3B

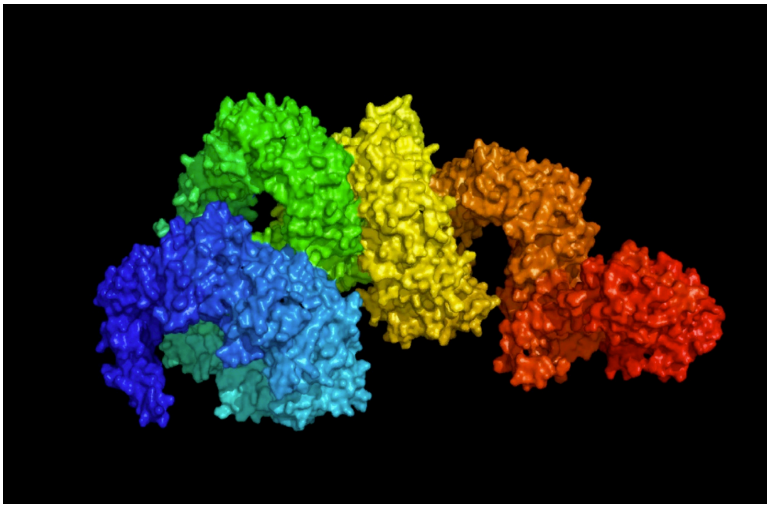
localization signal and a global negative hydropathy (-955) (figure 3A). Likewise, most known T4S effectors are hydrophilic, having total negative hydropathy scores on the Kyte-Doolittle scale (figure 3B). In addition, Erip1 harbors coiled-coil domains, which are protein domains commonly found in virulence effectors that interact with DNA, and shows a strong C-terminal basicity, which is a feature found in all known *alpha-proteobacteria* T4S effectors.

According to our analyses, Erip1 has several EPIYA like-motifs in its sequence. When we looked at the genomic localization of Erip1 we noticed that this potential T4S effector is located in a region with high G+C% and low gene density within the genome (data not shown). Taken together, these results suggested that Erip1 is an effector of *E. ruminantium*. Interestingly, we realized a BLAST analysis and found no-homolog in other bacteria present in public databases, suggesting that this protein may complete a function very specific to the developmental cycle of *E. ruminantium.* Consequently, we did a structure based function annotation of Erip1 protein using I-TASSER and COFACTOR (Zhang, 2008; A. Roy et al., 2012) and constructed a 3D model, which showed homology to a nucleoporin template with TM-score of 0.7 - 0.8 depending of the hit in PDB library (TM-score measures the global structural similarity between query and template protein) (figure 4A). Erip1 is predicted to have a helicoidal structure and revealed a large diameter, defined as the maximum distance between the coordinates of atom pairs, of approximately 230 Angström. NLS Mapper software confirmed the central localization of the NLS within the predicted structure (figure 4B).

Erip1 targets to the host cell nucleus

We identified a nuclear localization signal in the central part of Erip1 based on two *in silico* prediction programs: NLS Mapper software and S4TE algorithm (Kosugi et al., 2009; D. F. Meyer et al., 2013). Therefore, we examined whether the secreted Erip1 targets the nucleus during infection of BAE cells using immunofluorescence labeling. We stained the DNA with SYBR Green. Monoclonal anti-Erip1 and Alexa Fluor 647-conjugated goat anti-rabbit IgG were used for Erip1 labeling. In the early developmental cycle, Erip1 was localized in the host cytoplasm at 24 hpi when protein was the most abundant (figure 5). At 48 hpi, Erip1 localized in the nucleus

A



B

Figure 4

and cytoplasm of host cells (figure 5). Finally, Erip1 was detected in the cytoplasm at 72 hpi, corresponding at the stationary phase but the protein levels decreased.

Potential tyrosine phosphorylation of Erip1 EPIYA-related motifs

Bioinformatic analysis of Erip1 protein sequence showed the presence of tandemly repeated and duplicated sequences (figure 6A). The C-terminal portion contained three EPIYA-like motifs with the sequence EGPYR embedded in a conserved 17-amino-acid segment rich in proline residues (P) and which may be phosphorylated. Tyrosine phosphorylation of T4SS effectors may be important in order to hijack host cell functions (Böhmer et al., 2013). Unlike *A. phagocytophilum* AnkA protein, which has 4 different EPIYA motifs (IJdo et al., 2007), Erip1 harbors the same EPIYA-repeated sequences, resembling *H. pylori* CagA (Selbach et al., 2002). Additionally, we detected seven positively charged residues within the last 25 amino acids in the C-terminal, which could serve as signal sequences as seen in other T4SS effectors (Christie, 2004) and DUF3514 domains (figure 6A). We used a monoclonal antibody against phosphorylated tyrosines in Western Blots to determine if the EPIYA-like motifs within Erip1 are phosphorylated during infection of BAE cells with *E. ruminantium*. Western blot analyses showed one phosphorylated protein at 460 kDA in infected BAE cells and elementary bodies, which was not detected in uninfected BAE cells (figure 6B). This band corresponds with the molecular size detected by the monoclonal anti-Erip1 antibodies, suggesting that Erip1 is phosphorylated within its tyrosine residues.

E. ruminantium attenuated strain shows a different profile for the phosphorylation of Erip1

We used Western blots to determine the differences in the Erip1 profile in the attenuated strain compared to the virulent strain at different time points. When attenuated *E. ruminantium* was grown in BAE cells different versions of Erip1 were detected using the anti-Erip1 monoclonal antibodies. These versions included the full-length (460 kDa) and smaller cleaved versions (171, 117 and 14 kDa), which were expressed at 48, 72, 96 hpi. An additional difference in the protein expression between strains was when the peak expression occurred. Erip1 from the attenuated

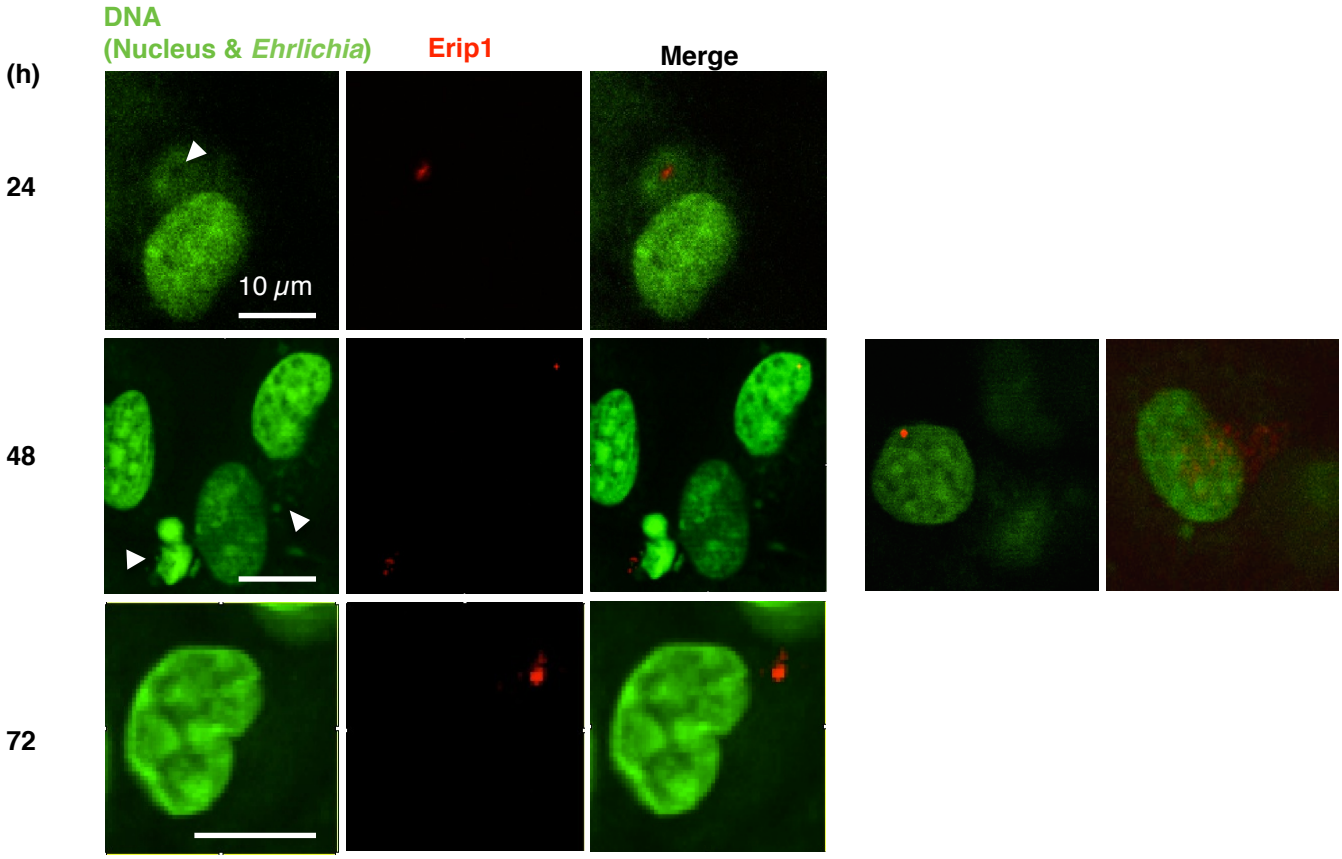


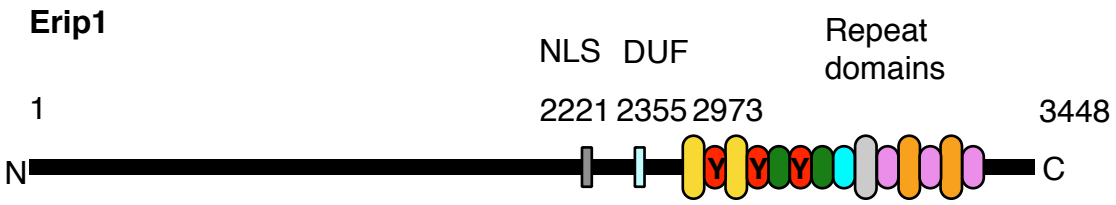
Figure 5

strain exhibited peaks at 24 and 48 hpi (figure 7A). Then, we investigated Erip1 phosphorylation at each of the time points in the attenuated strain. It is interesting to point out that only Erip1 cleaved versions were tyrosine phosphorylated (figure 7B).

Discussion

Bacterial effector proteins play crucial role in pathogenicity. In the present study, we demonstrated that ERGA_CDS_00570 was translocated in T4SS-dependent manner by *L. pneumophila*. Furthermore, a 100 amino acids portion of the C-terminal was sufficient for the translocation of this protein by the Dot/Icm system of *L. pneumophila*. The T4SS from *E. ruminantium* is more similar to that of *Agrobacterium tumefaciens virB/D4* system (Collins et al., 2005), which is different from that of *L. pneumophila*. However, the *L. pneumophila* translocation system has been previously used to screen the secretion of *A. marginale* T4SS effectors (Lockwood et al., 2011). This system was used to demonstrate the secretion of the *A. marginale* AnkA homolog, thus demonstrating the recognition of some of the signal peptides present in the *Anaplasmataceae* family (Lockwood et al., 2011). However, not all AnkA-like proteins seem to be translocated by this system. A different protein, which contained ANK domains (AnkC) and was predicted to be an effector in *A. marginale*, was not also translocated by *L. pneumophila* T4SS (Lockwood et al., 2011). Similarly, ERGA_CDS_03830, a homolog of AnkA from *A. phagocytophilum*, scored well in the S4TE algorithm but was not translocated by *L. pneumophila.* Therefore, we cannot exclude the possibility that the five other putative effectors of *E. ruminantium* are in fact translocated by T4SS, but did not possess the correct signal peptides for translocation by the *L. pneumophila* system. Erip1 is the largest protein encoded within *E. ruminantium*'s genome. It has a length of 3448 amino acids (460 kDa) and according to our BLAST results, it does not show homology to any other known bacterial protein, suggesting that it may complete a specific function unique to *E. ruminantium*. The C-terminal contains basic amino acids, similar to several T4SS effectors identified in *A. tumefaciens* T4S substrates, and harbors a total hydropathy score of -955, criteria found in many known effectors (Vergunst et al., 2005). A further characterization of the translocation signal within the different *E. ruminantium's* effectors is needed, since this particular signal peptide may be different to the majority of known effectors in *Anaplasmataceae*.

A



GGTDYLASTNSGVFNFQESSADEGSSEMLMKHYDQSLEHRQHSEYRLFDSTHQDSVDDGNIFSLRRNNFDEIS
NRVVGGRFENLPDTRMDFSYEERLRNLSLHQGDGYNDVALASNDVLGLRQEGPYRMDRPVVGGRFENLPDT
RMDFSYEERLRNLSLHQGDGYNDVALASNNVRSLPQEGPYRMDRPVAGAFARNLSGTQLDQSSDRDVALAS
SNVRSLPQEGPYRIDRPVAGAFARNLSGTQLDQSSDRDVALASNYVLSPSMVSDLITQVFARNLQRNRVDHSS
YEYRVHDPLLHHSSSRNIALASNDVPSLQQNSPDRVNNPDLPSSYMHSSREYRVHSPLLYQSGDSNVASVSDN
MSSLQQNSPDRVGNPDLPSSYMHSSREYRVHNPLLYQSDDSNVASVSDNMSSLQQNSPGRVDSLFVEGCTPSL
EHGLVGGHDELHNPLLCEKIGDNSFSSSVRSRGPRFYKEKSHQSPIGPRFCKARSHLTTHLASTELSDVDYQSG
QHPVTGNLIQNNPNNPSN**RR**S**R**IPLSLQNT**R**VS**K**TSS**RR**YVSN

Putative NLS

DUF3514 (pfam12027)

Y EPIYA-related EGPYR motif

40-aa L-rich repeat

17-aa PR-rich repeat

26-aa AS-rich repeat

17-aa spacer

Imperfect 39-aa S-rich repeat

10-aa SQ-rich repeat

39-aa S-rich repeat

B

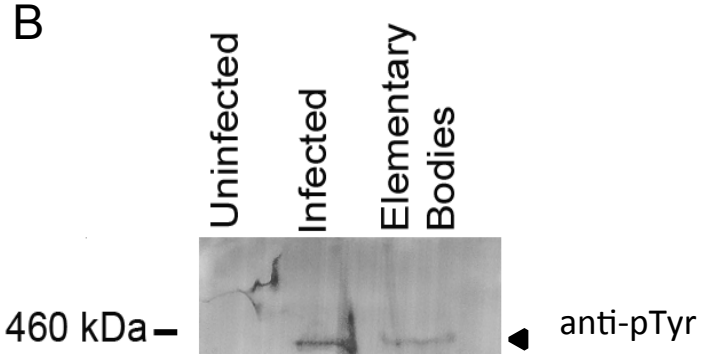
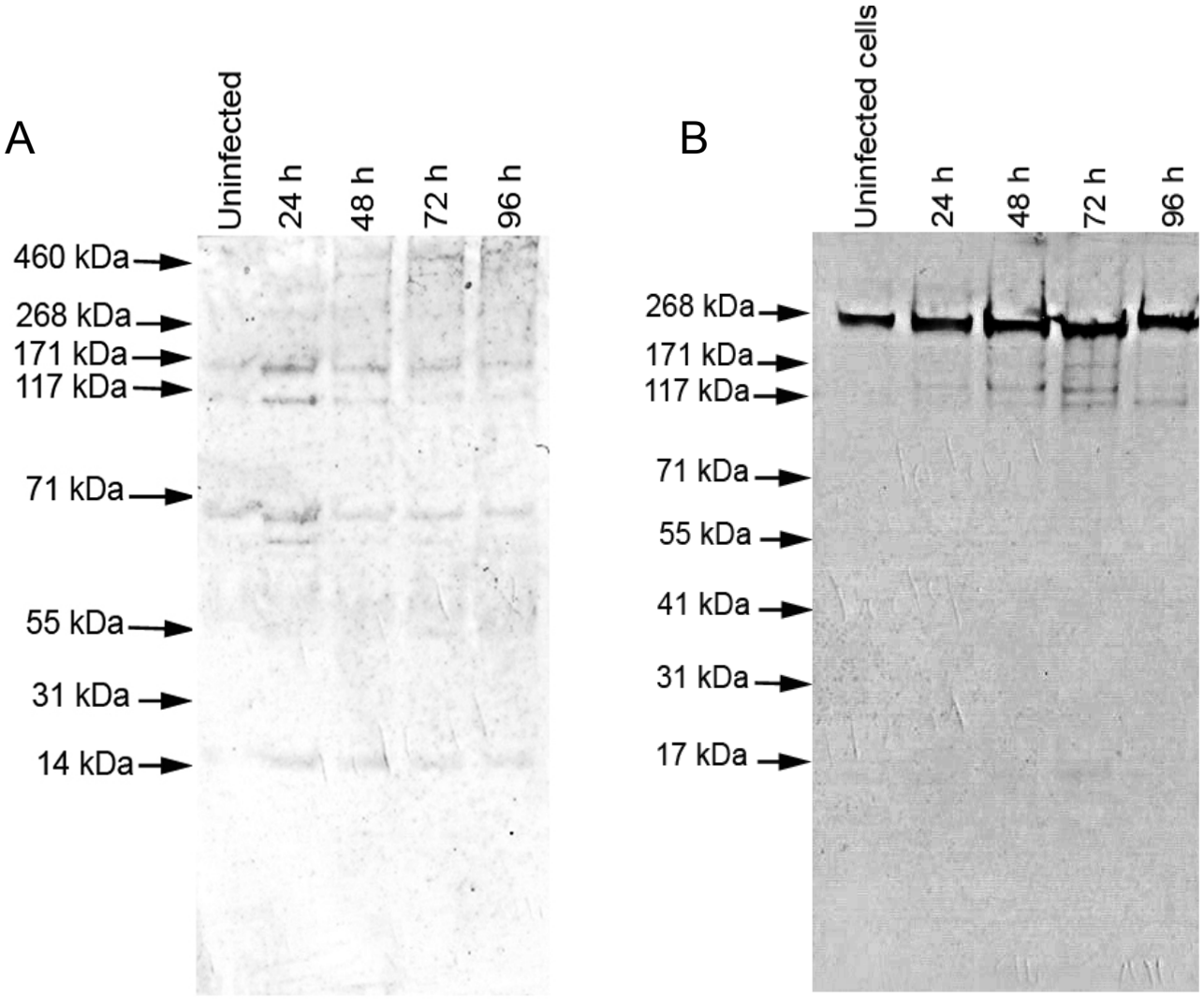


Figure 6

However, not many effectors have been identified within the *Anaplasmataceae* family. AnkA is the first effector identified in *A. phagocytophilum* and several members of the *Anaplasma* and *Ehrlichia* genera present homologs of this protein. AnkA is secreted by *A. phagocytophilum*'s T4SS and is tyrosine phosphorylated in the cytoplasm of the host cell from where is then transported into the nucleus. Once in the nucleus, AnkA interacts with the regulatory regions of the *CYBB* gene, resulting in the silencing of this and other genes involved in host defense (Garcia-Garcia et al., 2009; Rennoll-Bankert and Dumler, 2012). Only one additional effector has been identified in *A. phagocytophilum*, Ats-1, which localizes to the cytoplasm of infected cells where it sequesters autophagosomes by interacting with BECLIN1 (Niu et al., 2010). *E. chaffeensis* harbors an ortholog of Ats-1, ECH0825 which has 21% amino acid identity and localizes to the host cell mitochondria and inhibits apoptosis by inducing the expression of manganese superoxide dismutase (Liu et al., 2012). Other effectors have been identified in *E. chaffeensis* but they are not secreted by the T4SS, limiting the number of T4SS effectors to only six within the *Anaplasmataceae* family. This is in contrast to over 275 T4SS effectors that have been validated in *L. pneumophila* (Hubber and C. R. Roy, 2010).

According to our data, the Erip1 mRNA levels presented two peaks, one at 48 h and another at 120 hpi. However, according to the protein ratios the peaks of expression were at 24 h and 48 h during early phases of infection and the levels at 120 h showed the lowest ratio. It is possible that this discrepancy is due to the lost of the protein that is in the cytoplasm of infected cells during lysis at 120 hpi. A similar reduction in the protein levels compared to the mRNA levels was shown for the effector ECH0825 in *E. chaffeensis* (Liu et al., 2012). The early expression of Erip1 correlates with the peaks of expression of the T4S components in *E. chaffeensis* according with the results presented by Cheng *et al.* (Cheng et al., 2008) suggesting that once the T4S components were expressed and the apparatus assembled, Erip1 protein could be readily translocated to the cytoplasm of the host cell. Immunofluorescence assay indicated that Erip1 was translocated into the nuclei of endothelial cells during *E. ruminantium* infection. Bacterial proteins that target the nuclei of host cells may alter cell biology, which is a common pathogenic mechanism of bacteria (Bhavsar et al., 2007; Moon et al., 2012).



Relative ER input	1	0.98	1.57	2.95	14.78
Relative Erip1/ ER input	1	11.53	4.06	2.26	0.45

Relative ER input	1	0.98	1.57	2.95	14.78
Relative Erip1/ ER input	1	2.55	12.59	5.56	0.62

Figure 7

It has been shown that the expression of the *virBD* genes from the *E. chaffeensis* T4SS is driven by a transcription factor named EcxR and that the expression of the genes is different depending of the development stage of the bacteria (Cheng et al., 2008). *E. ruminantium* harbors a homolog of EcxR, that might also regulates Erip1 and its secretion. To verify this regulation, gel shift assays can be used to show binding of putative regulator to regions upstream of the *erip1* gene

In order to best characterize the hypothetical protein Erip1, a 3D structure was generated using the I-TASSER server (Zhang, 2008). This is an online platform used for the prediction of protein structures and functions. 3D models are built based as described by Zhang (Zhang, 2008). According to the results from the predictions, Erip1 may act as a nucleomodulin. However, the specific function of this protein and what proteins it interacts with remains as an answered question.

Some specific effectors are targeted to the nucleus and modulate or alter the immune response of the host. One of such pathogens is *Shigella* spp., which secretes OspF and $IpaH_{9.8}$ into the host cell then targeting it to the nucleus. In the nucleus, OspF dephosphorylates mitogen-activated protein kinases required for the transcription of NF-κB regulated genes, whereas $IpaH_{9.}$ interacts with a splicing factor that is involved in the expression of inflammatory cytokines (Zurawski et al., 2009). Likewise, *Salmonella enterica serovar typhimurium* harbors a leucine-rich repeat effector protein that inhibits NF-κB-dependent gene expression (Haraga and Miller, 2003). Other effectors that interact with defense genes have been identified in several pathogens, some examples include NUE in *Chlamydia trachomatis* (Pennini et al., 2010), YopM in *Yersinia species* (Benabdillah et al., 2004), EspF in *E. coli*, (Nougayrede and Donnenberg, 2004), and the recently identified LntA in *L. monocytogenes* (Rohde, 2011). These proteins form part of a newly categorized group named nucleomodulins, which consist of nucleus targeted proteins that alter the transcription of host genes by interacting with the host cell chromatin (Bierne and Cossart, 2012). Only a few number of nucleomodulins have been characterized in *Anaplasmataceae*, including AnkA of *Anaplasma phagocytophilum*, as well as Ank200 and several tandem-repeat containing proteins from *Ehrlichia chaffeensis* (Garcia-Garcia et al., 2009; McBride, 2011; Luo and McBride, 2012; Dunphy et al., 2013). Whether or not Erip1 belongs to this family of proteins, remains to be

experimentally determined, however, ChIP-on-chip can be used to investigate if Erip1 targets DNA in the host cells. Functional analysis could further determine if the DNA regions targeted by Erip1 are associated with transcriptional regulation, apoptosis, or vesicle trafficking *in vivo* (Pruneau et al., 2012). Likewise, the determinants for the localization of Erip1 can be study more deeply. The different putative NLSs that we identified in the present study can be studied by transfecting endothelial cells with plasmid constructs containing GFP-NLS fused proteins generated with the Gateway cloning system. The localization of GFP-tagged NLS proteins can be determined by confocal microscopy and compared with GFP clones lacking the NLS signal (Lee et al., 2012).

Furthermore, in a search for host proteins that can interact with Erip1 protein, we can conduct a yeast two-hybrid screening of bovine endothelial cell cDNA library using full length Erip1 as bait. This strategy was used previously in *Ehrlichia chaffeensis* to identify the host proteins interacting with a tandem repeat containing protein translocated by the T1SS and then modified by host proteins that add the small ubiquitin-like modifier (SUMO) (Dunphy et al., 2014). This protein interacts with a big array of host proteins and influences and wide range of host cell functions from transcription to cytoskeletal organization and it is essential for intracellular development of *E. chaffeensis* (Dunphy et al., 2014).

We found that Erip1 underwent tyrosine phosphorylation when *E. ruminantium* infected endothelial cells (figure 8). Phosphorylation of translocated effector proteins by host tyrosine kinases is a conserved mechanism within various bacterial infections. Many of these effectors are phosphorylated within motif containing repeated sequences of the amino acids EPIYA or modifications of this sequence, EPIYA-related motifs (Backert and Selbach, 2005). Our bioinformatics analysis showed three EPIYA related-motifs at the C-terminal of Erip1. We confirmed the phosphorylation of Erip1 in both the virulent and attenuated strains of *E. ruminantium.* However, the kinase involved in the phosphorylation of Erip1 remains unknown. Two different types of protein modules, termed Src homology 2 (SH2) and phosphotyrosine binding domains, can bind to tyrosines in a phosphorylation specific manner (Moran et al., 1990). AnkA from *A. phagocytophilum* is phosphorylated by Abl-1 tyrosine kinase and by Src kinases. This phosphorylation allows the

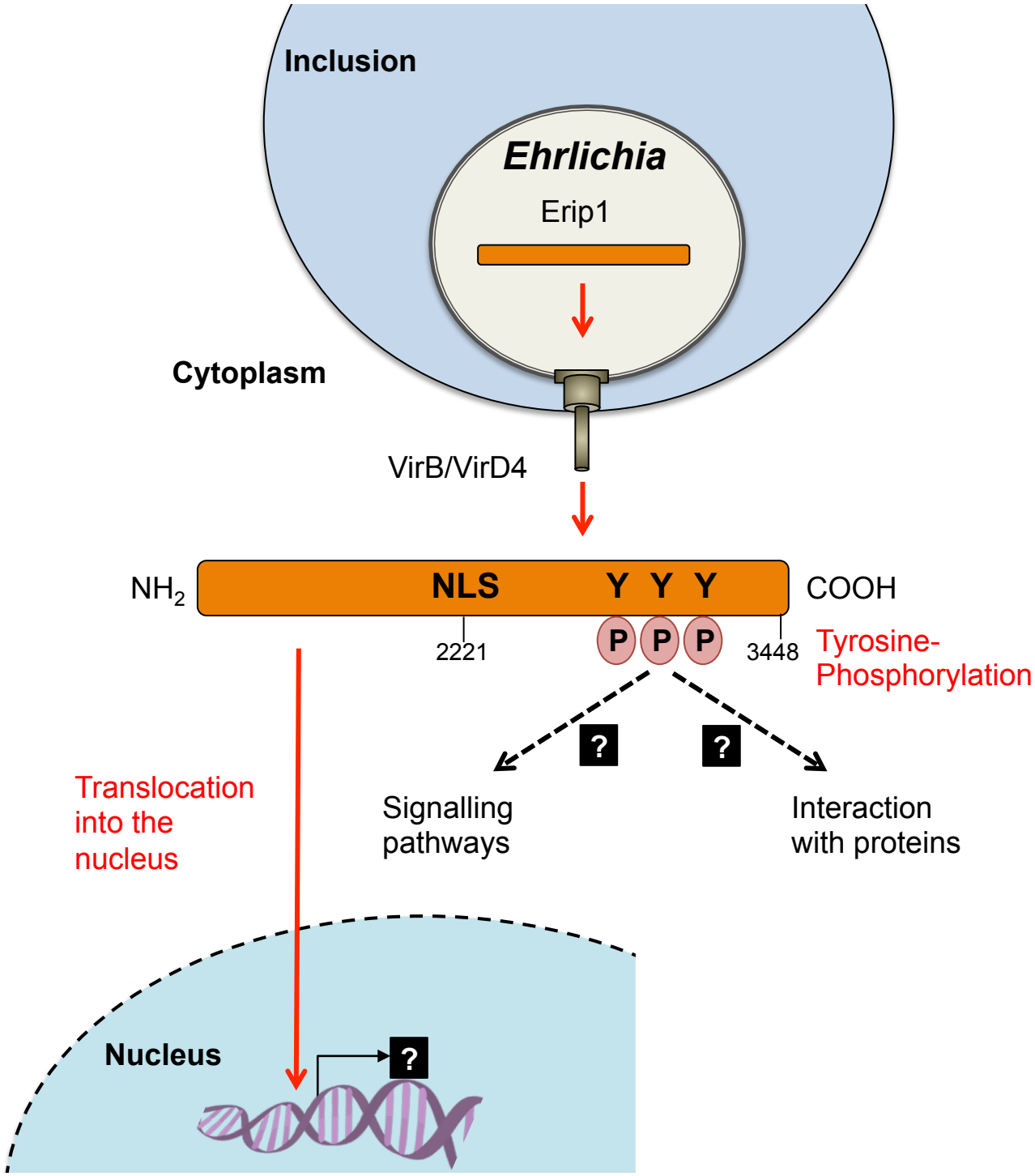

Inclusion
Ehrlichia
Erip1
Cytoplasm
VirB/VirD4
NH2
NLS
Y Y Y
COOH
2221
P P P
3448
Tyrosine-
Phosphorylation
?
?
Translocation
into the
nucleus
Signalling
pathways
Interaction
with proteins
Nucleus
?

Figure 8

subsequent binding of the SH2 domain from Shp-1 and affects the regulation of genes within the host nucleus (Rikihisa et al., 2010). Another example is Tir from *E. coli*, which is phosphorylated, binds to the SH2 and SH3 containing protein Nck and affects the assembly of actin filaments and pedestal formation (Rohatgi et al., 2001).

We have shown that Erip1 is cleaved in the attenuated strain and that only the cleaved parts are phosphorylated. It would be interesting to determine if the cleavage of the protein occurred pre- or post-phosphorylation. It is possible that Erip1 was cleaved by proteases in the attenuated strain due to the lack of chaperone protection. Whether or not this cleavage affects all the functions of Erip1 remains to be determined. But, it is likely that some of the functions completed by this protein are affected by its cleavage. Attenuated strains of *E. ruminantium* have a faster developmental cycle and appear to be more fitted for *in vitro* growth, even though they cannot infect mammalian host upon challenge. Thus, it is possible that the functions completed by the entire Erip1 are not necessary for *in vitro* infection and it is just required upon inoculation *in vivo*.

In conclusion, our data presented here demonstrate a novel T4E in the *Anaplasmataceae* family that is unique to *E. ruminantium* and shows no similarity with other known proteins. We demonstrated that Erip1 is tyrosine-phosphorylated and targeted into host cell nucleus. Our bioinformatics analysis suggests that Erip1 may be a nucleomodulin that could interfere with host cell signaling pathway. However, functional analyses need to be done to determine the exact function of this protein. The discovery of host cell partners will help us define the function of Erip1 and use it as antimicrobial biological effector for therapies (Rüter and Hardwidge, 2013).

Acknowledgments

We are grateful to M. Adoue at C3MAG for its help with confocal microscopy. We thank Y. Zhang for assistance to determine the predictive 3D structure of Erip1.

The authors acknowledge financial support from FEDER grant FED 1/1.4-30305, 2007–2013, “Risque en santé animale et végétale” (PhD grant to AM).

Table and Figure legends

Table 1. Oligonucleotide primers used for this study

Gene	Primer		Target size (bp)	Study
	Direction	Sequence (5'-3')		
erip1	F	GCTGAGACTGTTTCATTGTCATATTCTAC	273	quantitative RT-PCR
	R	CAGAAAGAACATAACCATTTAATATATA		
	c100aa-F	GCGGATCCTAAACAACCTTATAAGAGCG	300	Transclocation assay
	R	AGGTCGACTCAGTTTGATACATAGCGTCTTG		

Figure 1. CyaA translocation assays. Intracellular cAMP levels were determined following infection of THP-1 cells with *L. pneumophila* expressing CyaA fused to individual *E. ruminantium* protein. Results are expressed as fold change over cAMP levels resulting from infection with *L. pneumophila* expressing CyaA alone (negative control). Increased cAMP levels were observed when C-terminal of Erip1, AmAnkA (positive-control AM705 of *A. marginale*) fusion proteins were expressed in wild-type *L. pneumophila*, and levels similar to the negative-control were observed following expression of protein in DotA-deficient *L. pneumophila*, indicating that secretion requires a functional Dot/Icm T4SS. Results are shown for one experiment performed in six replicates and are representative of at least 2 individual experiments. One asterisk, $P < 0.01$ (as determined by Student's *t* test).

Figure 2. *E. ruminantium* growth and temporal expression of *erip1*.
A. Western blot analysis of Erip1, *E. ruminantium*-infected endothelial cells, purified elementary bodies, uninfected endothelial cells using rabbit anti-Erip1. B. Synchronous growth of *E. ruminantium* determined by quantitative PCR. Genomic DNA extracted from infected BAE bovine cells at different times p.i. was subjected to real-time PCR analysis. The data indicate the numbers of bacteria relative to the number at 0 hpi. C. Temporal expression of *erip1* in *E. ruminantium*-infected BAE cells as determined by quantitative real-time RT-PCR. Transcript amounts were normalized to the *E. ruminantium recA* gene. D. Temporal expression of Erip1 in *E. ruminantium*-infected BAE cells. Protein samples were subjected to Western blot

analysis based on bacteria number as determined by quantitative PCR. ER input: relative ratios of *E. ruminantium* loaded in SDS-PAGE wells.

Figure 3A. Analysis of T4S effector features for Erip1 (ERGA_CDS_00570) using S4TE software. Only positive hits are shown and corresponding domains or features were highlighted. The sequences of identified Nuclear Localization Signals, E-block and Coiled-coil domains are also indicated.

Figure 3B. Analysis of T4S effector features for Erip1 (ERGA_CDS_00570) using S4TE software. The total hydrophobicity of amino acid residues is represented in a Kyte-Doolittle scale. The more positive is in yellow and the less positive is in blue.

Figure 4. I-TASSER prediction of the 3D structure of Erip1 of *E. ruminantium* (Zhang, 2008). A. The 3D model has an helicoidal shape. B. Color scheme: α helices, β sheets and loops are in blue, pink and orange colour respectively. The nuclear localization signal is represented in red. PyMOL was applied to subsequent graphics processing.

Figure 5. Translocation of Erip1 from *E. ruminantium* into the cytoplasm and nucleus of endothelial cells. SYBR Green was used to label DNA (green). *E. ruminantium*-infected cells were subjected to immunofluorescence labeling using rabbit anti-Erip1 (Erip1; AF647, red) at different times p.i. Merge, merged images. Arrows indicate the bacteria. Scale bar: 10 µm

Figure 6. Analysis of the C-terminal sequence of Erip1

A. Schematic of Erip1 and the C-terminal amino acid sequence. The C-terminal contains a series of repeated sequences. There are three EPIYA-like motifs consisting of EPIYA-related EGPYR motif. Each EPIYA-like motif is part of a 17-amino-acid repeat sequence with proline-rich region. The positively charged amino acids (in bold at the C-terminus) are characteristic of the T4SS signal sequence.

B. Erip1 is phosphorylated in infected BAE cells. Immunoblot of lysates of uninfected and infected BAE cells was probed with anti-phosphotyrosine (anti-pTyr) antibodies.

Figure 7. Temporal expression of Erip1

A. Immunoblot of lysates of uninfected and infected BAE cells with attenuated *E. ruminantium* strain during developmental cycle was probed with anti-Erip1 antibodies. B. Erip1 is phosphorylated in infected BAE cells. Immunoblot of lysates of uninfected and infected BAE cells with attenuated *E. ruminantium* strain was probed with anti-phosphotyrosine (anti-pTyr) antibodies.

Figure 8. Subcellular localization of Erip1 and host cell signaling pathways. *Ehrlichia* type IV secretion effector Erip1 is translocated into the host cell cytoplasm. It contains three tyrosine phosphorylation sites. Erip1 is phosphorylated in the cytoplasm, where signaling pathways or interaction with other proteins can occur. Erip1 protein translocates into the nucleus and may execute gene regulation

Supplementary figure 1: Western blot showing expression of *E. ruminantium* proteins as CyaA fusions in *Legionella*.

Supplementary figure 2: CyaA translocation assays of the 6 T4E candidates.

References

Allsopp, B.A. (2010) Natural history of *Ehrlichia ruminantium*. *Veterinary Parasitology* **167**: 123–135.

Backert, S. and Meyer, T.F. (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. *Current Opinion in Microbiology* **9**: 207–217.

Backert, S. and Selbach, M. (2005) Tyrosine-phosphorylated bacterial effector proteins: the enemies within. *Trends in Microbiology* **13**: 476–484.

Barré, N., Uilenberg, G., Morel, P., and Camus, E. (1987) Danger of introducing heartwater on to the American mainland : potential role of indigenous and exotic *Amblyomma* ticks **. 54**: 405–417.

Benabdillah, R., Jaime Mota, L., Lützelschwab, S., Demoinet, E., and Cornelis, G.R. (2004) Identification of a nuclear targeting signal in YopM from *Yersinia* spp. *Microb. Pathog.* **36**: 247–261.

Bhavsar, A.P., Guttman, J.A., and Finlay, B.B. (2007) Manipulation of host-cell pathways by bacterial pathogens. *Nature* **449**: 827–834.

Bierne, H. and Cossart, P. (2012) When bacteria target the nucleus: the emerging family of nucleomodulins. *Cell Microbiol* **14**: 622–633.

Böhmer, F., Szedlacsek, S., Tabernero, L., Östman, A., and Hertog, den, J. (2013) Protein tyrosine phosphatase structure-function relationships in regulation and pathogenesis. *FEBS J.* **280**: 413–431.

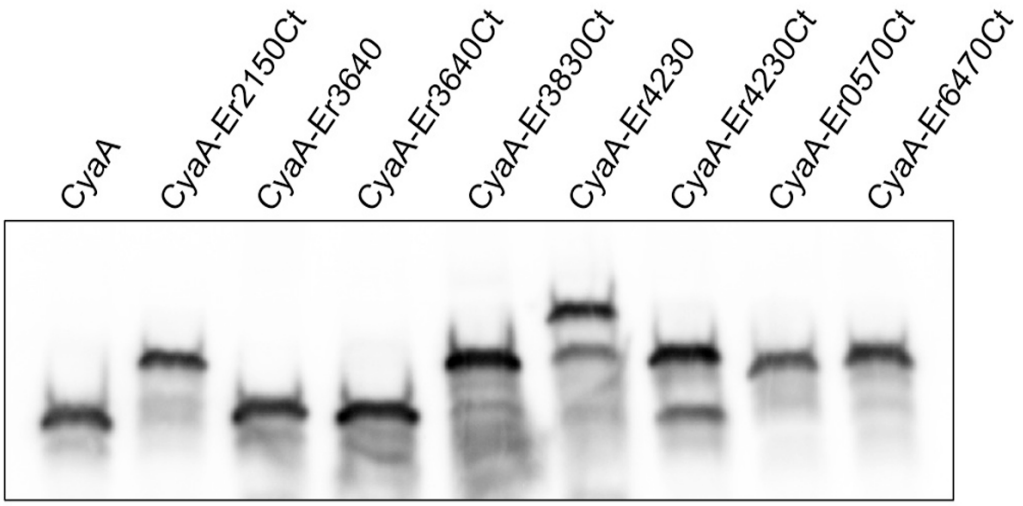
Campellone, K.G., Giese, A., Tipper, D.J., and Leong, J.M. (2002) A tyrosine-phosphorylated 12-amino-acid sequence of enteropathogenic *Escherichia coli* Tir binds the host adaptor protein Nck and is required for Nck localization to actin

pedestals. *Molecular Microbiology* **43**: 1227–1241.
Cascales, E. and Christie, P.J. (2003) The versatile bacterial type IV secretion systems. *Nat Rev Micro* **1**: 137–149.
Cheng, Z., Wang, X., and Rikihisa, Y. (2008) Regulation of type IV secretion apparatus genes during *Ehrlichia chaffeensis* intracellular development by a previously unidentified protein. *Journal of Bacteriology* **190**: 2096–2105.
Christie, P.J. (2004) Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems. *Biochim. Biophys. Acta* **1694**: 219–234.
Clifton, D.R., Fields, K.A., Grieshaber, S.S., Dooley, C.A., Fischer, E.R., Mead, D.J., et al. (2004) A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 10166–10171.
Collins, N.E., Liebenberg, J., de Villiers, E.P., Brayton, K.A., Louw, E., Pretorius, A., et al. (2005) The genome of the heartwater agent Ehrlichia ruminantium contains multiple tandem repeats of actively variable copy number. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 838–843.
Deng, K., Mock, J.R., Greenberg, S., van Oers, N.S.C., and Hansen, E.J. (2008) *Haemophilus ducreyi* LspA proteins are tyrosine phosphorylated by macrophage-encoded protein tyrosine kinases. *Infect. Immun.* **76**: 4692–4702.
Deng, W., Vallance, B.A., Li, Y., Puente, J.L., and Finlay, B.B. (2003) *Citrobacter rodentium* translocated intimin receptor (Tir) is an essential virulence factor needed for actin condensation, intestinal colonization and colonic hyperplasia in mice. *Molecular Microbiology* **48**: 95–115.
Dunning Hotopp, J.C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S.V., Eisen, J.A., et al. (2006) Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet* **2**: e21.
Dunphy, P.S., Luo, T., and McBride, J.W. (2014) *Ehrlichia chaffeensis* exploits host SUMOylation pathways to mediate effector-host interactions and promote intracellular survival. *Infect. Immun.* **82**: 4154–4168.
Dunphy, P.S., Luo, T., and McBride, J.W. (2013) *Ehrlichia* moonlighting effectors and interkingdom interactions with the mononuclear phagocyte. *Microbes Infect.* **15**: 1005–1016.
Emboulé, L., Daigle, F., Meyer, D.F., Mari, B., Pinarello, V., Sheikboudou, C., et al. (2009) Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium*. *BMC Mol. Biol.* **10**: 111.
Garcia-Garcia, J.C., Rennoll-Bankert, K.E., Pelly, S., Milstone, A.M., and Dumler, J.S. (2009) Silencing of host cell *CYBB* gene expression by the nuclear effector AnkA of the intracellular pathogen *Anaplasma phagocytophilum*. *Infect. Immun.* **77**: 2385–2391.
Haraga, A. and Miller, S.I. (2003) A *Salmonella enterica* serovar typhimurium translocated leucine-rich repeat effector protein inhibits NF-kappa B-dependent gene expression. *Infect. Immun.* **71**: 4052–4058.
Hatakeyama, M. (2004) *Helicobacter pylori* causes gastric cancer by hijacking cell growth signaling. *Discovery Medicine* **4**: 476–481.
Hubber, A. and Roy, C.R. (2010) Modulation of Host Cell Function by *Legionella pneumophila*Type IV Effectors. *Annu. Rev. Cell Dev. Biol.* **26**: 261–283.
IJdo, J.W., Carlson, A.C., and Kennedy, E.L. (2007) *Anaplasma phagocytophilum*

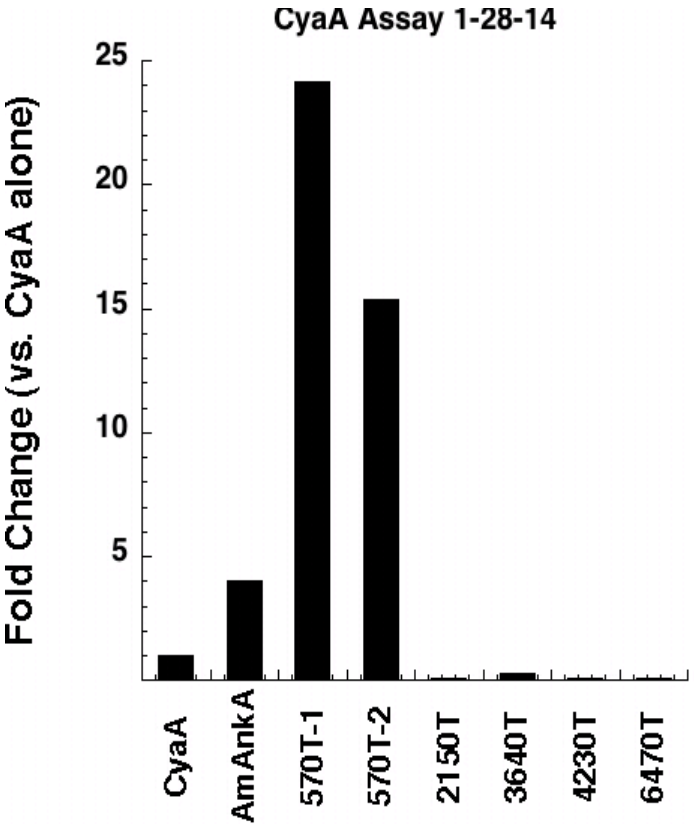
AnkA is tyrosine-phosphorylated at EPIYA motifs and recruits SHP-1 during early infection. *Cell Microbiol* **9**: 1284–1296.
Jongejan, F., Zandbergen, T.A., van de Wiel, P.A., de Groot M, and Uilenberg, G. (1991) The tick-borne rickettsia Cowdria ruminantium has a Chlamydia-like developmental cycle. *The onderstepoort journal of veterinary research* **58**: 227–237.
Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U.S.A.* **106**: 10171–10176.
Lee, J.H., Jun, S.H., Baik, S.C., Kim, D.R., Park, J.-Y., Lee, Y.S., et al. (2012) Prediction and screening of nuclear targeting proteins with nuclear localization signals in Helicobacter pylori. *Journal of Microbiological Methods* **91**: 490–496.
Liu, H., Bao, W., Lin, M., Niu, H., and Rikihisa, Y. (2012) *Ehrlichia* type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating host MnSOD. *Cell Microbiol* **14**: 1037–1050.
Lockwood, S., Voth, D.E., Brayton, K.A., Beare, P.A., Brown, W.C., Heinzen, R.A., and Broschat, S.L. (2011) Identification of *Anaplasma marginale* type IV secretion system effector proteins. *PLoS ONE* **6**: e27724.
Luo, T. and McBride, J.W. (2012) *Ehrlichia chaffeensis* TRP32 interacts with host cell targets that influence intracellular survival. *Infect. Immun.* **80**: 2297–2306.
Marcelino, I., Veríssimo, C., Sousa, M.F.Q., Carrondo, M.J.T., and Alves, P.M. (2005) Characterization of *Ehrlichia ruminantium* replication and release kinetics in endothelial cell cultures. *Veterinary Microbiology* **110**: 87–96.
McBride, J.W. (2011) Ehrlichia chaffeensis tandem repeat proteins and Ank200 are type 1 secretion system substrates related to the repeats-in-toxin exoprotein family. 1–19.
Mehlitz, A., Banhart, S., Hess, S., Selbach, M., and Meyer, T.F. (2008) Complex kinase requirements for *Chlamydia trachomatis* Tarp phosphorylation. *FEMS Microbiology Letters* **289**: 233–240.
Meyer, D.F., Noroy, C., Moumène, A., Raffaele, S., Albina, E., and Vachiery, N. (2013) Searching algorithm for type IV secretion system effectors 1.0: a tool for predicting type IV effectors and exploring their genomic context. *Nucleic Acids Research* **41**: 9218–9229.
Moon, D.C., Gurung, M., Lee, J.H., Lee, Y.S., Choi, C.W., Kim, Il, S., and Lee, J.C. (2012) Screening of nuclear targeting proteins in *Acinetobacter baumannii* based on nuclear localization signals. *Research in Microbiologoy* **163**: 279–285.
Moran, M.F., Koch, C.A., Anderson, D., Ellis, C., England, L., Martin, G.S., and Pawson, T. (1990) Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 8622–8626.
Niu, H., Kozjak-Pavlovic, V., Rudel, T., and Rikihisa, Y. (2010) *Anaplasma phagocytophilum* Ats-1 is imported into host cell mitochondria and interferes with apoptosis induction. *PLoS Pathog* **6**: e1000774.
Nougayrede, J.-P. and Donnenberg, M.S. (2004) Enteropathogenic *Escherichia coli* EspF is targeted to mitochondria and is required to initiate the mitochondrial death pathway. *Cell Microbiol* **6**: 1097–1111.
Pan, X., Lührmann, A., Satoh, A., Laskowski-Arce, M.A., and Roy, C.R. (2008) Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors.

Science **320**: 1651–1654.
Pawson, T. and Scott, J.D. (2005) Protein phosphorylation in signaling--50 years and counting. *Trends in Biochemical Sciences* **30**: 286–290.
Pennini, M.E., Perrinet, S., Dautry-Varsat, A., and Subtil, A. (2010) Histone methylation by NUE, a novel nuclear effector of the intracellular pathogen *Chlamydia trachomatis. PLoS Pathog* **6**: e1000995.
Prozesky, L. and Plessis, Du, J. (1987) Heartwater. The development and life cycle of *Cowdria ruminantium* in the vertebrate host, ticks and cultured endothelial cells.*The onderstepoort journal of veterinary research* **54**: 193–196.
Pruneau, L., Emboulé, L., Gely, P., Marcelino, I., Mari, B., Pinarello, V., et al. (2012) Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development. *FEMS Immunol Med Microbiol* **64**: 66–73.
Rennoll-Bankert, K.E. and Dumler, J.S. (2012) Lessons from *Anaplasma phagocytophilum*: chromatin remodeling by bacterial effectors. *Infect Disord Drug Targets* **12**: 380–387.
Rikihisa, Y. and Lin, M. (2010) *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins. *Current Opinion in Microbiology* **13**: 59–66.
Rikihisa, Y., Lin, M., and Niu, H. (2010) Type IV secretion in the obligatory intracellular bacterium *Anaplasma phagocytophilum*. *Cell Microbiol* **12**: 1213–1221.
Rohatgi, R., Nollau, P., Ho, H.Y., Kirschner, M.W., and Mayer, B.J. (2001) Nck and phosphatidylinositol 4,5-bisphosphate synergistically activate actin polymerization through the N-WASP-Arp2/3 pathway. *Journal of Biological Chemistry* **276**: 26448–26452.
Rohde, J.R. (2011) *Listeria* unwinds host DNA. *Science* **331**: 1271–1272.
Roy, A., Yang, J., and Zhang, Y. (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research* **40**: W471–7.
Rüter, C. and Hardwidge, P.R. (2013) “Drugs from Bugs”: bacterial effector proteins as promising biological (immune-) therapeutics. *FEMS Microbiology Letters* **351**: 126–132.
Selbach, M., Moese, S., Hauck, C.R., Meyer, T.F., and Backert, S. (2002) Src is the kinase of the *Helicobacter pylori* CagA protein *in vitro* and *in vivo*. *Journal of Biological Chemistry* **277**: 6775–6778.
Selbach, M., Paul, F.E., Brandt, S., Guye, P., Daumke, O., Backert, S., et al. (2009) Host cell interactome of tyrosine-phosphorylated bacterial proteins. *Cell Host Microbe* **5**: 397–403.
Sutten, E.L., Norimine, J., Beare, P.A., Heinzen, R.A., Lopez, J.E., Morse, K., et al. (2010) *Anaplasma marginale* type IV secretion system proteins VirB2, VirB7, VirB11, and VirD4 are immunogenic components of a protective bacterial membrane vaccine. *Infect. Immun.* **78**: 1314–1325.
Vachiery, N., Marcelino, I., Martinez, D., and Lefrançois, T. (2013) Opportunities in diagnostic and vaccine approaches to mitigate potential heartwater spreading and impact on the American mainland. *Developments in biologicals* **135**: 191–200.
Vergunst, A.C., van Lier, M.C.M., Dulk-Ras, den, A., Stüve, T.A.G., Ouwehand, A., and Hooykaas, P.J.J. (2005) Positive charge is an important feature of the C-

terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium.* *Proc. Natl. Acad. Sci. U.S.A.* **102**: 832–837.
Voth, D.E., Broederdorf, L.J., and Graham, J.G. (2012) Bacterial Type IV secretion systems: versatile virulence machines. *Future Microbiology* **7**: 241–257.
Voth, D.E., Howe, D., Beare, P.A., Vogel, J.P., Unsworth, N., Samuel, J.E., and Heinzen, R.A. (2009) The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. *Journal of Bacteriology* **191**: 4232–4242.
Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40.
Zurawski, D.V., Mumy, K.L., Faherty, C.S., McCormick, B.A., and Maurelli, A.T. (2009) *Shigella flexneri* type III secretion system effectors OspB and OspF target the nucleus to downregulate the host inflammatory response via interactions with retinoblastoma protein. *Molecular Microbiology* **71**: 350–368.



Supplementary figure 1



Supplementary figure 2

DISCUSSION GENERALE

Discussion générale et perspectives

Les bactéries pathogènes ont développé une multitude de facteurs de virulence dans le but de manipuler leur hôte pour échapper à la réponse du système immunitaire et établir une niche réplicative, compartiment sûr dans lequel elles peuvent proliférer. Au cours du processus infectieux, les bactéries intracellulaires pathogènes doivent faire face à des conditions environnementales hostiles dans l'animal et aux mécanismes de défense induits par l'animal après perception du pathogène. Pour mettre en place la maladie, ces bactéries ont développé des systèmes de régulation fins leur permettant de coordonner, en fonction de stimuli environnementaux, l'expression des différents déterminants de virulence dont elles disposent. Le but principal de ma thèse a été d'identifier certains déterminants moléculaires du pouvoir pathogène chez *Ehrlichia ruminantium* ainsi que la régulation de leur expression en fonction de stimuli environnementaux ou la caractérisation de leur fonction. Dans ce but, j'ai mené plusieurs approches en parallèle qui visaient à i) caractériser sans *a priori* les protéines de la membrane externe d'*E. ruminantium*, ii) étudier la régulation de l'expression des gènes codant le système de sécrétion de type IV (SST4), les protéines de la famille Map1, le facteur de transcription *tr1* et le régulateur ErxR et iii) à identifier les substrats protéiques transitant par ce SST4 et analyser leur rôle dans la cellule hôte. Je vais discuter ci-dessous les résultats de mes travaux en les intégrant à l'ensemble des données de la littérature.

1. Protéines de la membrane externe, interface critique pour les interactions cellule hôte-bactérie pathogène

La première étape de l'interaction cellule hôte-bactérie pathogène est le contact physique entre la bactérie et la surface de la cellule hôte. Cette fonction est assurée par la membrane externe chez les bactéries à Gram négatif. Cette barrière physique protège les bactéries contre différents stress sans compromettre l'échange de matériel nécessaire à la survie des bactéries. Plus particulièrement, les protéines de la membrane externe jouent un certain nombre de rôles clés chez les bactéries leur permettant de s'adapter à différents environnements ainsi qu'à leurs hôtes. Parmi les diverses fonctions assurées par ces protéines, on trouve la biogénèse et l'intégrité de la membrane externe, des activités de porines non spécifiques, le transport actif, l'adhésion ou des activités enzymatiques associées aux membranes (Koßmehl et al. 2013). De nombreuses études sur les bactéries pathogènes ont montré un rôle important de ces protéines dans la stimulation de la réponse immune de l'hôte et dans la protection de celui-ci vis-à-vis d'une infection (Agnes et al. 2011).

Les bactéries du genre *Ehrlichia* sont des pathogènes intracellulaires obligatoires animaux et zoonotiques majeurs. Au cours de leur cycle de développement, ces bactéries

s'attachent et rentrent dans la cellule hôte eucaryote selon un processus d'endocytose médié par les protéines de la membrane externe (Mohan Kumar et al. 2013).

Au cours de cette étude, nous avons optimisé un protocole de purification des complexes protéiques de la membrane externe d'*E. ruminantium* par lyse chimique à l'aide du détergent sarkosyl. Nous avons ensuite analysé les fractions membranaires obtenues par chromatographie en phase liquide (nano-LC) couplée à la spectrométrie de masse (MALDI-TOF-TOF MS/MS), et identifié plusieurs protéines de la membrane externe.

Nous avons confirmé la présence des protéines membranaires connues dans cette fraction membranaire mais aussi celle de protéines jamais décrites expérimentalement comme faisant partie de la membrane externe. Nous avons également identifié quelques protéines normalement présentes dans les autres fractions (cytoplasme ou périplasme) mais pouvant être associées à la membrane externe. Les résultats de cette étude du protéome de la membrane montrent qu' *E. ruminantium* possèderait certains déterminants majeurs de virulence pour infecter les cellules hôtes lorsque la bactérie est sous forme infectieuse, c'est-à-dire au stade corps élémentaire. En effet, nous avons identifié certaines protéines de la famille Map1 (Major Antigenic Protein 1), des lipoprotéines Bam (Beta-barrel assembly machine) impliquées dans l'assemblage des protéines dans la membrane externe, certaines protéines du SST4 associées à la virulence et à la sécrétion et une protéine impliquée dans le transport de nutriments.

La protéine de la membrane externe la plus étudiée chez *E. ruminantium* est la protéine Map1 qui déclenche une réponse humorale dominante chez les animaux infectés par la bactérie (van Vliet et al. 1994). Notre analyse a permis d'identifier cinq protéines de la famille MAP1 dans la fraction membranaire, soit une de plus que ce qui avait été trouvé dans une étude antérieure sur le protéome total d'*E. ruminantium* (Marcelino et al. 2012). Ces protéines ont fait l'objet d'étude comme potentiels antigènes vaccinaux pour le développement de vaccins recombinants, notamment avec l'utilisation d'un vaccin ADN exprimant le gène *map1*, conférant une protection entre 23 et 88% chez la souris (Nyika et al. 1998). Cependant, la protéine Map1 ne procure pas une protection efficace lors d'essais vaccinaux sur ruminants (van Kleef, Neitz, and De Waal 1993). La faible efficacité de protection des protéines Map1 semble reposer sur le polymorphisme des protéines Map1 entre différentes souches d'*E. ruminantium* ou la présence d'isoformes pour une même protéine (Barbet, Byrom, and Mahan 2009; Marcelino et al. 2012). De plus, des études transcriptomiques ont mis en évidence un profil d'expression différent des gènes *map1* en fonction du stade de développement la bactérie mais aussi de l'hôte (animal ou vecteur). Ainsi, en utilisant une approche protéomique chez *A. marginale* isolées à partir d'érythrocytes, Noh *et al.* ont identifié des protéines de surface de la membrane externe d'*A. marginale* surexprimées (Noh et al. 2008). L'immunisation avec ces fractions membranaires

induit une protection efficace contre la maladie chez des animaux infectés par *A. marginale*. De plus, la comparaison du protéome de surface d'*A. marginale* isolé à partir d'érythrocytes ou de cellules de tiques a permis d'identifier certaines protéines spécifiquement exprimées dans les érythrocytes et une protéine de surface (AM778) spécifiquement exprimée chez la tique. Ainsi ce remodelage de la membrane externe d'*A. marginale* permettrait à la bactérie le passage des cellules hôtes mammifères à son hôte arthropode (Noh et al. 2008). Chez *E. ruminantium*, le gène *map1-1* n'est induit que dans les cellules de tiques et *map1-6* n'est exprimé que dans le corps réticulé (Postigo et al. 2007; Pruneau et al. 2012). Nous pourrions comparer le profil d'expression dans les cellules endothéliales et cellules de tiques afin d'identifier de nouvelles cibles pour bloquer la transmission.

Certaines protéines du complexe Bam ont été également identifiées (BamA et BamD). La protéine BamA est centrale dans ce complexe et est essentielle dans l'assemblage des protéines de la membrane externe (Knowles et al. 2009). Des études chez *Borrelia burgdorferi* ont montré qu'une mutation sur le gène *bamA* induit une forte diminution de l'intégration des protéines dans la membrane externe (Lenhart and Akins 2010). Cibler et inactiver cette protéine permettrait de rendre la membrane externe de la bactérie plus fragile, diminuant ainsi son rôle protecteur et son rôle dans la virulence. Dans notre étude, nous avons également détecté deux protéines du SST4 d'*E. ruminantium,* VirB9 et VirB10, connues pour être dans la membrane externe (Jakubowski et al. 2005). Comme nous l'avons déjà décrit de manière exhaustive, le SST4 permet d'injecter des effecteurs protéiques pour détourner la machinerie cellulaire de l'hôte pour l'acquisition de nutriments, leurrer les systèmes cellulaires de défense ou créer des compartiments intracellulaires propices à la survie et à la prolifération bactérienne (Voth, Broederdorf, and Graham 2012). Des études ont montré que certaines protéines du SST4 pouvaient être immunogènes et potentiellement conférer une protection du bétail infecté par *Anaplasma marginale* (Lopez et al. 2007). En effet, l'immunisation avec des fractions de la membrane externe purifiées de la bactérie a montré une protection chez les animaux infectés et a révélé un panel de protéines immunogènes dont VirB9 et VirB10. Certaines protéines du SST4 représentent donc de bons candidats vaccinaux contre la bactérie (Sutten et al. 2010).

Un des résultats marquants est l'identification des protéines membranaires hypothétiques et une nouvelle porine. Ces protéines potentiellement antigéniques pourraient être reconnues par les récepteurs de l'immunité innée et être capables d'éliciter des réponses anticorps neutralisantes *in vivo*. Il serait donc très intéressant de les tester pour leur pouvoir immunogène protecteur en conditions contrôlées.

Les porines permettent le passage d'ions et de petites molécules organiques dans les deux sens : de l'extérieur vers l'intérieur pour l'arrivée des nutriments, et de l'intérieur vers

l'extérieur pour la détoxification du cytoplasme (rejet de produits secondaires toxiques par accumulation) ou la sécrétion de protéines ou de molécules organiques (Galdiero et al. 2012). On pourrait envisager de bloquer cette porine pour empêcher l'acquisition de certains nutriments par la bactérie, ce qui pourrait freiner sa prolifération à l'intérieur de la cellule hôte. De même, cette porine pourrait être utilisée comme un potentiel candidat vaccinal et évaluée pour son pouvoir protecteur contre *Ehrlichia*.

De plus, il serait intéressant de comparer le profil d'expression des protéines de la membrane externe de la souche Gardel virulente et la souche atténuée cultivées en laboratoire comme ce qui a été déjà réalisé chez *Rickettsia prowazekii* (Bechah et al. 2010). Les études transcriptomique et protéomique ont mis en évidence des différences majeures concernant certaines protéines de surface Sca entre les souches virulente et atténuée. Certains gènes impliqués dans la biogénèse de la membrane sont sous-exprimés chez la virulente par rapport à l'atténuée. De plus, des mutations sur ces gènes affecteraient leur fonction dans la virulence parmi les différentes souches de *R. prowazekii*. Ceci suggère que la souche virulente *R. prowazekii* n'exprime pas certaines de ses protéines de surface pour échapper à la reconnaissance du système immunitaire.

La mutagénèse des bactéries intracellulaires obligatoires telles que *E. chaffeensis* ou *A. phagocytophilum* a été récemment mise au point (C. Cheng et al. 2013; Felsheim et al. 2010). Pour identifier les facteurs de virulence critiques chez *E. ruminantium*, nous envisageons de générer par mutagénèse aléatoire par transposon Himar une banque de mutants marqués à la GFP et à la luciférase. Cette banque sera ensuite criblée à haut-débit pour les phénotypes liés à la pathogenèse, à l'instar de la stratégie utilisée chez *Coxiella burnetii* où plus de 3000 mutants ont été isolés et 1082 annotés et séquencés (Martinez et al. 2014). Plusieurs facteurs régulant les différentes étapes d'infection de la bactérie ont ainsi été identifiés tels que ceux impliqués dans l'internalisation de la bactérie à l'intérieur de la cellule hôte, la biogénèse de la vacuole ou la réplication intracellulaire, la protection des cellules infectées contre l'apoptose. Surtout, cette étude a permis d'identifier la première invasine OmpA chez *C. burnetii*. Une mutation sur ce gène induisant une diminution significative de l'internalisation et réplication de la bactérie, soulignant ici l'importance des protéines de la membrane externe dans l'interaction hôte-pathogène (Martinez et al. 2014).

2. Régulation de l'expression des gènes codant le SST4 et les protéines de la famille Map1 chez *E. ruminantium* : perception de l'environnement par la bactérie

Pour parvenir à infecter avec succès son hôte animal, la bactérie doit réguler très finement l'ensemble des fonctions liées à la pathogenèse de manière à adapter

temporellement, spatialement et quantitativement l'expression génique aux modifications des conditions environnementales.

Des études antérieures ont montré que le SST4 n'était pas exprimé de façon constitutive mais finement régulé par des facteurs de transcription (Altman and Segal 2008; De Jong et al. 2008; Martínez-Núñez et al. 2010; Xueqi Wang, Kikuchi, and Rikihisa 2007; Zusman et al. 2007). Le facteur de transcription EcxR, identifié chez *E. chaffeensis*, se lie aux promoteurs en amont des gènes *virBD* du SST4 qu'il induit durant le cycle de développement de la bactérie (Z. Cheng, Wang, and Rikihisa 2008). De plus, ce régulateur présente une boucle de rétrocontrôle positif. L'orthologue de ce gène, *apxR*, est également un facteur de transcription retrouvé chez *A. phagocytophilum* (Xueqi Wang, Kikuchi, and Rikihisa 2007). Cette protéine s'autorégule également mais régule aussi positivement le facteur de transcription *tr1* et les gènes codant pour les protéines de surface de la famille P44 (Xueqi Wang et al. 2007; Xueqi Wang, Kikuchi, and Rikihisa 2007). Ces protéines de la membrane externe P44 engendrent une forte réponse humorale chez les patients malades ou les souris infectées expérimentalement (Ijdo et al. 1998; Zhi et al. 2002). Ce sont des porines qui facilitent le transport de sucres ou d'acides aminés à travers la membrane externe de la bactérie (Huang et al. 2007). Une étude récente a montré la présence de facteurs de transcription homologues régulant positivement le SST4 chez *Wolbachia*, une autre *Anaplasmataceae* (Li and Carlow 2012). Comme attendu, il existe un orthologue chez *E. ruminantium* et nous avons démontré qu'il s'autorégulait et régulait également positivement les gènes du SST4. Il apparaît donc que ce mode de régulation soit assez bien conservé chez les *Anaplasmataceae*. Ces facteurs de transcription qui régulent l'expression des gènes du SST4 sont attractifs pour le développement d'antibiotiques (Baron and Coombes 2007). Nous avons également montré que la protéine ErxR active également le facteur de transcription *tr1* (homologue de celui d'*A. phagocytophilum*) en amont des gènes du cluster *map1* en se liant à son promoteur. *tr1* étant le promoteur principal des protéines de surface de la famille P44 chez *A. phagocytophilum* (Barbet et al. 2005), il est probable que *tr1* régule l'expression des gènes de la famille *map1* de manière similaire chez *E. ruminantium*. La fonction de ce gène reste à ce jour inconnue mais comme ce qui a été montré chez *A. phagocytophilum*, il se pourrait que *tr1* ait un rôle dans l'infection de la tique puisque ce gène est surexprimé dans les cellules de tique, contrairement à *apxr* qui lui est exprimé seulement dans les cellules humaines (Nelson et al. 2008). En revanche, nous n'avons pas pu démontrer qu'ErxR régulait directement les gènes *map1*. Ce résultat suggère que la régulation par ErxR n'est peut-être pas directe (régulation par *tr1*) ou qu'il peut exister un autre régulateur intervenant sur les gènes *map1* et permettant de coupler/intégrer les différents signaux environnementaux. En effet, chez *Brucella abortus*, l'expression des gènes *virB* du SST4 est sous le contrôle de divers facteurs de transcription qui permettent à ce système de répondre aux différents types de

signaux de l'environnement tels que l'acidification ou des carences en nutriments (Arocena, Zorreguieta, and Sieira 2012; Sieira et al. 2012; Viadas et al. 2010). Des expériences de gel retard, chromatographie d'affinité et spectrométrie de masse ont permis de mettre en évidence des protéines se liant au promoteur des gènes *virB* de *Brucella* comme la protéine IHF, intervenant dans la modification de structure d'ADN avant acidification (Porte, Liautard, and Köhler 1999). Il serait donc intéressant de rechercher aussi la protéine régulatrice des gènes *map1* chez *E. ruminantium*.

En plus d'être régulés par ce régulateur ErxR, nous avons démontré que les gènes du SST4 et les gènes codant pour les protéines de la famille *map1* sont induits en réponse à des signaux environnementaux comme l'acidité ou la carence en fer.

Certaines bactéries comme *C. burnetii,* sont capables de survivre et se multiplier dans une vacuole acide à l'intérieur de la cellule hôte (Romano et al. 2007). L'acidité de la cellule hôte est d'ailleurs un signal d'activation du système de sécrétion de type III (SST3) pour d'autres bactéries telles que *Salmonella* spp. ou *Edwardsiella ictaluri* (Steele-Mortimer 2008). Ainsi, le SST3 de *Salmonella typhimurium* est induit par l'acidification du pH de la cellule, consécutive à la reconnaissance de la bactérie par des récepteurs TLR de l'immunité (toll-like receptors) de la cellule hôte (Arpaia et al. 2011). Ce résultat est surprenant puisque ces TLRs contribuent à la résistance contre les agents pathogènes. Cependant, les auteurs ont montré que chez des souris transgéniques n'exprimant plus les TLR 2 et 4, la bactérie ne pouvait plus exprimer le SST3, ni se répliquer. Il semblerait donc que la bactérie ne déclenche la pathogenèse qu'à la condition d'être dans sa cellule hôte, le signal ultime de reconnaissance de l'hôte étant l'acidification résultant de l'activation des TLRs suite à l'infection de la cellule (Arpaia et al. 2011). Dans notre étude, nous avons supposé que la vacuole dans laquelle se développait *E. ruminantium* était acide d'après les données présentes chez *E. chaffeensis* où il a été démontré que l'endosome dans lequel se réplique la bactérie avait un pH acide (Z. Cheng, Lin, and Rikihisa 2014). Nous avons donc confirmé que la vacuole d'*E. ruminantium* était acide et puis nous avons voulu mimer les conditions *in vivo* en appliquant un stress acide au stade corps élémentaire lorsque la bactérie est à l'extérieur de la cellule. Nous avons montré que les gènes du SST4, les gènes *map1* et *erxR* étaient bien exprimés en condition de pH acide.

Nous avons également démontré que la carence en fer affectait l'expression des gènes codant pour les déterminants du pouvoir pathogène que nous avons mentionné plus haut. Le fer est un cofacteur essentiel pour de nombreuses réactions enzymatiques, mais il est peu soluble et en quantités infinitésimales sous forme libre dans l'environnement (Skaar 2010). Pour satisfaire leur besoin en fer, les bactéries ont développé de nombreux systèmes de captation du Fe^{3+} extrêmement efficaces pour solubiliser et séquestrer ce métal aux concentrations micromolaires nécessaires à leur métabolisme (Skaar 2010; Teixidó et al.

2011). En intervenant dans le métabolisme des bactéries, le fer a donc un rôle indirect dans le pouvoir pathogène, mais il peut aussi être un déterminant direct de la virulence (Xin Wang et al. 2009). Ainsi, chez *Moraxella catarrhalis*, des mutants déficients dans la production d'hème ont un défaut de croissance important (de Vries et al. 2013). La carence en fer chez l'hôte infecté est un également un signal environnemental important pour l'expression des gènes de virulence du SST3 de *Bordetella* (Brickman et al. 2011). Chez *E. ruminantium*, nous avons montré que la carence en fer induisait également l'expression du SST4, de certains gènes *map* et du régulateur *erxR*. La carence en fer serait donc un signal environnemental important pour l'expression du SST4 chez *E. ruminantium*. Ces observations mettent en relation la régulation de la virulence d'*E. ruminantium* et les stimuli environnementaux et nutritionnels spécifiques de l'hôte. Cette première étude nous confirme le lien étroit entre la carence en fer, l'acidité et la virulence d'*E. ruminantium*. L'ensemble de ces réponses mettent en exergue la façon dont à *E. ruminantium* perçoit son environnement, détecte qu'elle est dans sa cellule hôte et active les mécanismes de pathogenèse nécessaire à sa croissance.

D'autres approches plus globales ont été entreprises chez d'autres bactéries comme la comparaison de profils d'expression de gènes de *Bacillus anthracis* en condition de carence ou supplémentation en fer (P. E. Carlson et al. 2009). Récemment, une approche globale couplant analyses transcriptomiques et fonctionnelles a permis d'identifier le gène *iroT* de *Legionella pneumophila* et de montrer que son induction en condition de carence en fer est liée à la virulence de la bactérie (Portier et al. 2014). La mutation de ce gène a en effet un impact négatif sur la croissance intracellulaire de la bactérie. C'est une protéine de membrane qui serait donc impliquée dans le transport du fer. Nous pourrions envisager d'adopter la même stratégie chez *E. ruminantium* afin d'identifier les gènes clés dans l'acquisition du fer et dans la croissance de la bactérie. Au cours cette étude, nous avons également vu que les gènes *map1* étaient également exprimés suite à la perception de la carence en fer et que ces gènes codent pour des porines qui seraient donc susceptibles de laisser passer des nutriments comme le fer. Ce rôle nous rappelle celui des récepteurs TonB-dépendants qui sont des protéines de la membrane externe connues pour intervenir dans le transport des complexes Fer-sidérophores chez les bactéries à Gram négatif (Krewulak and Vogel 2011). Ces récepteurs possèdent des boucles extracellulaires permettant la perception des signaux extérieurs (Ferguson and Deisenhofer 2004). Il est à noter que les protéines Map1 possèdent la même structure en tonneaux constitués de feuillets ß avec des boucles extracellulaires. Il est donc tentant de faire l'analogie fonctionnelle entre ces deux groupes de protéines et d'imaginer que les protéines Map1 pourraient donc intervenir dans la perception du fer. Comme décrit plus haut, nous avons vu que les gènes *map1* n'étaient pas régulés directement par Erxr, mais qu'ils étaient en revanche induits en condition de carence

en fer. Il serait donc intéressant de rechercher un autre régulateur de ces gènes *map1*. Chez de les bactéries, le métabolisme du fer est régulé par le répresseur global, Fur, qui se lie en présence de Fer au niveau des promoteurs des gènes régulés au niveau d'un motif particulier, la « Fur box » (Escolar, Pérez-Martín, and de Lorenzo 1998). En condition de carence en fer, la répression par Fur est levée, permettant ainsi la transcription des gènes impliqués dans la capture du fer. Nous avons identifié des motifs Fur putatifs en amont de plusieurs gènes *map* et il serait intéressant de voir si un homologue de cette protéine Fur existe chez *E. ruminantium* et s'il peut réguler les gènes *map1*, confirmant ainsi le rôle des gènes *map1* dans l'acquisition du fer.

3. Le SST4: « VIP pass » pour l'intérieur de la cellule hôte

Le système de sécrétion de type IV (SST4) présent chez de nombreuses bactéries intracellulaires agit comme une seringue moléculaire permettant l'injection active d'effecteurs protéiques de type IV (ET4) dans la cellule hôte. Ces ET4 servent à manipuler la machinerie cellulaire de l'hôte au profit de la bactérie et à déjouer les réponses immunitaires pour permettre la multiplication bactérienne.

3.1. Erip1, premier effecteur du SST4 d'*E. ruminantium*… et effecteur unique dans le règne bactérien

Nous avons développé S4TE (*Searching algorithm for type IV secretion system effectors*), un outil bioinformatique pour la prédiction d'ET4s basé sur la combinaison de 13 caractéristiques d'effecteurs connus (D. F. Meyer et al. 2013). Ce logiciel a permis d'identifier 22 effecteurs candidats chez *E. ruminantium*. Si certains étaient attendus, en raison de leurs homologies avec des effecteurs déjà caractérisés, d'autres semblent en revanche, être complètement nouveaux et spécifiques à *E. ruminantium*. Nous avons choisi plusieurs candidats pour valider leur translocation par le SST4 en système hétérologue chez *L. pneumophila* en utilisant le système CyaA (Lockwood et al. 2011). Ce criblage a permis l'identification d'un nouvel effecteur du SST4 d' *E. ruminantium*, qui a été nommé Erip1 pour *Ehrlichia Ruminantium Injected Protein* 1. Ce travail a été effectué en collaboration avec l'équipe de D. Voth (University of Arkansas for Medical Sciences). Cet effecteur unique dans le règne bactérien possède deux caractéristiques majeures: un domaine NLS et des motifs EPIYA-like répétés. Des expériences d'immunofluorescence en microscopie confocale ont confirmé qu'Erip1 est transféré dans le noyau des cellules endothéliales bovines. Cette observation nous laisse supposer qu'Erip1 pourrait appartenir à une nouvelle famille de nucléomodulines, effecteurs identifiés chez de nombreuses bactéries telles que *E. chaffeensis*, *Listeria monocytogenes*, *Xanthomonas campestris*, *A. phagocytophilum* (Bierne

and Cossart 2012). Ces nucléomodulines sont des effecteurs capables de pénétrer dans le noyau des cellules et de manipuler l'expression de gènes notamment en dérégulant la transcription ou la maturation de l'ARN. Il serait à présent intéressant d'identifier les cibles nucléaires d'Erip1. Des expériences d'immunoprecipitation par la chromatine (ChIP) ont permis d'identifier les cibles nucléaires de AnkA, effecteur d'*A. phagocytophilum* (Garcia-Garcia et al. 2009). Les auteurs ont montré que AnkA, qui est transloqué dans le noyau de la cellule hôte, interagit avec des gènes régulant la chromatine et inhibe la transcription du gène *CYBB* (*gp91^{phox}*) qui code pour un composant du complexe phagocyte NADPH oxidase et d'autres gènes clés de défense de la cellule hôte (Garcia-Garcia et al. 2009).

De plus, nous avons montré qu'Erip1 était phosphorylé au niveau des tyrosines chez la souche Gardel virulente et atténuée. La question que l'on peut se poser est de savoir si ces tyrosines sont phosphorylées au niveau des motifs EPIYA-like situés au niveau du C-terminal d'*erip1*. Pour cela, il s'agira de cloner la séquence EPIYA-like dans un vecteur d'expression, purifier la protéine, la soumettre à la kinase Src et vérifier si on a une phosphorylation (IJdo, Carlson, and Kennedy 2007). De plus, il est intéressant de noter que chez la souche atténuée, Erip1 semble clivé et n'est phosphorylé qu'au niveau de ces fragments clivés et non sur la forme entière. Il est donc possible qu'en raison de sa taille, Erip1 soit sécrété sous forme linéaire, protégé par une ou des protéines chaperonnes contre les dégradations d'éventuelles protéases. Il est possible que de telles chaperonnes ne soient plus exprimées dans la souche atténuée, supposant des niveaux de régulations différents dans les deux souches, ce qui aboutirait à la dégradation d'Erip1 dès sa translocation dans le cytoplasme de la cellule hôte. Ainsi, nous pourrions rechercher les interacteurs protéiques bactériens ou de la cellule hôte d'Erip1 en réalisant des expériences de pull down avec Erip1 couplé à des billes magnétiques et incubé avec de l'extrait de protéines totales bactériennes ou de la cellule hôte. Les interacteurs issus de cette expérience pourront être ensuite identifiés par spectrométrie de masse.

3.2. Et les autres effecteurs ?

Afin d'identifier d'autres effecteurs d'*E. ruminantium*, l'étude de la sécrétion par le SST4 d'autres candidats identifiés par S4TE est en cours en collaboration avec l'équipe de D. Voth. Nous avons choisi plusieurs effecteurs en fonction de certaines caractéristiques importantes connues et compilées par S4TE, comme l'hydrophobicité globale, la charge positive en C-terminal ou la présence de signaux de localisation. En parallèle de cette recherche d'effecteurs par méthode bioinformatique, nous avons entrepris la recherche d'effecteurs par une méthode biologique. Nous avons commencé à générer une banque d'ADN génomique d'*E. ruminantium* afin de cribler cette banque pour les effecteurs par un système

double hybride bactérien utilisant la protéine VirD4 comme appât. Cette méthode a notamment permis d'identifier l'effecteur ECH0825 chez *E. chaffeensis* et Ats-1 chez *A. phagocytophilum* (Liu et al. 2012; Niu et al. 2010).

3.3. Etude fonctionnelle des ET4s

Les fonctions de ces effecteurs putatifs pourraient être analysées en détail par des approches de biologie moléculaire ou cellulaire. En utilisant la microscopie confocale, comme pour Erip1 avec des anticorps spécifiques, nous pouvons confirmer la sécrétion des effecteurs *in vitro* et caractériser leur localisation subcellulaire à l'intérieur de la cellule hôte. Nous pourrons tester la cytotoxicité de ces effecteurs chez la levure mais surtout des expériences de gain ou perte de fonction dans les cellules endothéliales nous permettront de caractériser la fonction de ces effecteurs selon les phénotypes obtenus.

Enfin, nous avons vu plus haut que la mutagénèse aléatoire par le système Himar et dirigée par recombinaison homologue est maintenant possible chez les *Ehrlichia* (C. Cheng et al. 2013). Pour déterminer le rôle de chaque effecteur, il faudra analyser les phénotypes des mutants d'*E. ruminantium* dans ces effecteurs identifiés, lors les différentes étapes d'infection des cellules endothéliales bovines ou pour leur développement *in vivo* chez la souris.

3.4. Effectome de type IV, déterminant de la spécificité d'hôte *d'E. ruminantium*?

En raison de la co-évolution étroite avec leur réservoir d'hôtes, les bactéries pathogènes ont généralement limité le spectre d'hôtes qu'elles peuvent infecter avec succès. La spécificité d'hôte est un concept fondamental qui décrit la nature des associations entre la bactérie et l'hôte. Ce concept est largement étudié puisqu'il permet de comprendre les lois qui régissent la virulence potentielle de la bactérie pathogène lorsqu'elle envahit un nouvel environnement, hôte ou tissus, les zoonoses, les maladies émergentes et ré-émergentes.

Les changements dans la spécificité d'hôte de pathogènes animaux, en particulier, sont le sujet d'une préoccupation majeure dû à leur impact immédiat et inattendu sur la santé humaine. Le saut d'hôte peut être attribué à la modification de facteurs du pouvoir pathogène clés qui facilitent la mise en place d'une association particulière avec l'hôte (Kirzinger and Stavrinides 2012).

Les bactéries du genre *Ehrlichia* représentent d'excellents modèles d'étude pour identifier les facteurs bactériens impliqués dans la spécificité d'hôtes puisque chaque espèce possède un spectre d'hôte restreint *in vivo*. *E. ruminantium* infecte les ruminants en parasitant les cellules endothéliales mais certaines souches d'*E. ruminantium* sont également

capables d'infecter la souris. *E. ruminantium* est référencée comme agent pathogène animal malgré trois cas humains mortels répertoriés en Afrique du Sud mais controversés en raison de l'absence de confirmation par cytologie ou isolement de la bactérie (Allsopp, Louw, and Meyer 2005). Pourtant, *E. ruminantium* est capable d'infecter les cellules endothéliales humaines *in vitro*.

Il est donc légitime de se demander pourquoi *E. ruminantium* n'est pas un agent pathogène humain alors qu'elle peut infecter les cellules humaines *in vitro*. Pour répondre à cette question, il serait intéressant de comparer les répertoires d'effecteurs du SST4 entre différentes souches d'*E. ruminantium* et de déterminer quels sont les effecteurs spécifiques de chaque souche pouvant être liés à la spécificité d'hôte. Nous pourrions ainsi déterminer la distribution d'effecteurs potentiels chez différentes souches d'*E. ruminantium* par des amplifications de gènes spécifiques sur la base du répertoire d'effecteurs caractérisé dans la souche Gardel.

De plus, certaines *Ehrlichia* connues d'abord pour être des pathogènes vétérinaires sont capables d'infecter l'homme. C'est le cas d' *E. chaffeensis* et *E. ewingii* qui causent une importante maladie zoonotique émergente, l'ehrlichiose monocytique humaine (Rikihisa 2010). *E. chaffeensis* est également pathogène sur souris avec des niveaux de virulence différents selon les souches. Il serait donc pertinent de comparer les répertoires d'effecteurs candidats de différentes souches *Ehrlichia* pathogènes sur souris *(chaffeensis, ruminantium, muris)* afin d'identifier les effecteurs potentiellement caractéristiques des souches infectants les mammifères et éventuellement l'homme. Les résultats de ces analyses révèleront des répertoires d'effecteurs conservés et accessoires qui ont probablement des rôles distincts dans le pouvoir pathogène bactérien et des histoires évolutives différentes (Hajri et al. 2009).

Enfin, à plus long terme, il serait intéressant de caractériser le répertoire minimal d'effecteurs responsable de la virulence d'*E. ruminantium*. A l'instar du système PRIVAS développé chez la bactérie phytopathogène *Pseudomonas syringae* par Cunnac *et al.*, nous pourrions mettre en place des expériences de génétique combinatoire (Cunnac et al. 2011). Il s'agirait tout d'abord de muter l'ensemble des effecteurs majeurs du SST4 identifiés chez *E. ruminantium*. La perte totale de virulence de la bactérie serait alors vérifiée lors de tests pathogènes sur souris. Chaque effecteur serait ensuite cloné individuellement ou en combinaison avec d'autres effecteurs et nous pourrions réaliser des expériences de complémentation avec les clones portant les différentes combinaisons jusqu'à trouver le répertoire minimal d'effecteurs restaurant le phénotype virulent (Cunnac et al. 2011). Ceci serait très riche d'enseignement quand à l'évaluation de la redondance d'action entre effecteurs, leur dynamique de sécrétion mais permettrait aussi de mieux comprendre l'évolution de ces protéines en caractérisant leur rôle individuel dans le pouvoir pathogène d' *E. ruminantium*. Enfin, un tel projet mettrait en évidence les mécanismes cruciaux de la

cellule hôte ciblés par la bactérie et fournirait des informations fondamentales sur le fonctionnement cellulaire en réponse à une infection par une bactérie intracellulaire obligatoire.

Perspectives

La complémentarité des différentes approches menées au cours de cette thèse a permis de mieux comprendre les différentes stratégies mises en place par la bactérie pour infecter et se développer à l'intérieur de sa cellule hôte. L'analyse des déterminants moléculaires du pouvoir pathogène d'*E. ruminantium* a été conduite à des différents niveaux allant de l'étude de la régulation des gènes du SST4 et des gènes *map1* par des stimuli environnementaux au cours du cycle de développement, à l''étude d'un nouvel effecteur du SST4, en passant par l'identification des protéines de la membrane externe. L'ensemble des résultats obtenus permet d'avoir une vision plus globale dans l'espace et dans le temps de la mise en œuvre du pouvoir pathogène par la bactérie. La poursuite de ces travaux permettra d'imaginer un modèle dynamique du comportement de la bactérie au cours de son cycle de développement.

Le processus d'invasion par les bactéries intracellulaires consiste généralement en plusieurs étapes telles que l'adhésion, l'internalisation, la prolifération intracellulaire et la diffusion intercellulaire. Notre étude a permis d'identifier chez *E. ruminantium* une partie des mécanismes de régulation qui dictent l'expression de certains gènes de virulence permettant à la bactérie de proliférer et se développer à l'intérieur de la cellule hôte. Afin d'identifier des gènes éventuellement régulés par ErxR, nous pourrions réaliser une analyse comparative par RNA-seq du transcriptome d'*E. ruminantium* sauvage à celui d'*E. ruminantium* muté pour le gène *erxR*. L'analyse des phénotypes de mutants dans les gènes différentiellement exprimés dans la souche mutée permettrait de caractériser l'ensemble de la cascade de régulation dépendant d'ErxR mais aussi de voir si ErxR est impliqué dans d'autres voies de signalisation.

Cependant, il serait intéressant de voir quels sont les autres circuits de régulation gouvernant le pouvoir pathogène d'*E. ruminantium* et son adaptation à l'hôte. Ainsi, un autre mode de régulation bactérien a récemment été mis en évidence par l'identification de transferts horizontaux d'ADN non codant (Oren et al. 2014). En réponse à des signaux environnementaux, les bactéries semblent ainsi capables acquérir de nouvelles séquences régulatrices qui permettraient de réguler les protéines déjà présentes, favorisant une plus grande diversification de phénotypes infectieux (Oren et al. 2014). Le RNA silencing joue un

rôle majeur dans les défenses antibactériennes chez les plantes, animaux ou insectes grâce à l'action des microRNA (miRNA). Cependant certaines bactéries sont capables d'exploiter ces miRNA à leur avantage pour déstabiliser les défenses de l'hôte. Ainsi, *P. syringae* injecte plusieurs effecteurs du SST3 dans sa cellule hôte pour inhiber la transcription ou l'activité de certains miRNA (Navarro et al. 2008). Ces voies de régulation seraient un nouveau moyen pour les bactéries de remodeler leur environnement cellulaire afin qu'elles puissent se développer et s'adapter (Hakimi and Cannella 2011). Ces miRNA seraient donc de nouveaux régulateurs, que l'on pourrait identifier par des expériences de RNA-seq dans les cellules hôtes infectées.

De même, des approches de type RNA-seq permettraient d'avoir accès à un niveau de résolution supérieur pour l'étude de la régulation génétique dépendante des ARN chez *Ehrlichia*. Nous pourrions ainsi mettre à jour l'existence de nouveaux petits ARN régulateurs ainsi que des mécanismes insoupçonnés de régulation de la virulence d'*E. ruminantium* à l'instar de ce qui a été découvert chez *Listeria monocytogenes* (Cossart and Lebreton 2014). En effet, ces miRNA revêtent une importance croissante dans la régulation de la virulence des bactéries pathogènes intracellulaires et peuvent permettre une réponse rapide et versatile via de nouveaux mécanismes d'intégration des signaux environnementaux (Mellin et al. 2014). Il est également certain que nous pourrions découvrir de nouveaux déterminants de virulence par l'existence d'interférences entre les mécanismes de régulation liés aux ARNs de la bactérie et aux ARNs de l'hôte (Hakimi and Cannella 2011).

Ainsi, en raison des énormes progrès accomplis dans le domaine des nouvelles technologies de séquençage, d'analyse à haut-débit de l'expression des ARN et des protéines, de l'imagerie *in vivo* ou encore de la manipulation génétique des *Anaplasmataceae*, de nouvelles perspectives sont à présent envisageables pour étudier la réponse au stress environnementaux, la virulence et la transmission d'*E. ruminantium*. De nouveaux concepts sont générés tant au niveau de la régulation génique que de la communication cellulaire au sein des communautés bactériennes.

Le champ d'investigation immense qui s'ouvre devra privilégier une approche systémique afin de mieux appréhender la fonction des gènes, la détection des différents niveaux de régulation ainsi que les comparaisons entre espèces différentes. Ces connaissances générées permettront sans nul doute de proposer des alternatives thérapeutiques et de nouvelles stratégies de contrôle des maladies infectieuses (Rüter and Hardwidge 2013). Dans cette nouvelle ère de la Microbiologie, le modèle *Ehrlichia* reste pertinent et peut même être avantageux pour étudier des processus complexes comme la virulence bactérienne ou la perception de l'environnement par la bactérie.

REFERENCES BIBLIOGRAPHIQUES

Adakal, Hassane et al. 2010. "Efficiency of Inactivated Vaccines Against Heartwater in Burkina Faso: Impact of *Ehrlichia Ruminantium* Genetic Diversity.." *Vaccine* 28(29): 4573–80.

Agnes, Joseph T et al. 2011. "Identification of *Anaplasma Marginale* Outer Membrane Protein Antigens Conserved Between a. Marginale Sensu Stricto Strains and the Live *A. Marginale* Subsp. *Centrale* Vaccine.." *Infection and immunity* 79(3): 1311–18.

Allsopp, MT, M Louw, and E C Meyer. 2005. "*Ehrlichia Ruminantium*: An Emerging Human Pathogen?." *Annals of the New York Academy of Sciences* 1063: 358–60.

Altman, Efrat, and Gil Segal. 2008. "The Response Regulator CpxR Directly Regulates Expression of Several *Legionella Pneumophila Icm/Dot* Components as Well as New Translocated Substrates.." *Journal of Bacteriology* 190(6): 1985–96.

Arocena, Gastón M, Angeles Zorreguieta, and Rodrigo Sieira. 2012. "Expression of VjbR Under Nutrient Limitation Conditions Is Regulated at the Post-Transcriptional Level by Specific Acidic pH Values and Urocanic Acid.." *PLoS ONE* 7(4): e35394.

Arpaia, Nicholas et al. 2011. "TLR Signaling Is Required for *Salmonella Typhimurium* Virulence.." *Cell* 144(5): 675–88.

Barbet, Anthony F et al. 2005. "Identification of Functional Promoters in the Msp2 Expression Loci of *Anaplasma Marginale* And *Anaplasma Phagocytophilum*.." *Gene* 353(1): 89–97.

Barbet, Anthony F, Barbara Byrom, and Suman M Mahan. 2009. "Diversity of *Ehrlichia Ruminantium* Major Antigenic Protein 1-2 in Field Isolates and Infected Sheep.." *Infection and immunity* 77(6): 2304–10.

Barnewall, R E, Y Rikihisa, and E H Lee. 1997. "*Ehrlichia Chaffeensis* Inclusions Are Early Endosomes Which Selectively Accumulate Transferrin Receptor.." *Infection and immunity* 65(4): 1455–61.

Baron, Christian, and Brian Coombes. 2007. "Targeting Bacterial Secretion Systems: Benefits of Disarmament in the Microcosm.." *Infectious disorders drug targets* 7(1): 19–27.

Baumgartner, Wes A et al. 2014. "Modulation of Vacuolar pH Is Required for Replication of *Edwardsiella Ictaluri* In Channel Catfish Macrophages.." *Infection and immunity* 82(6): 2329–36.

Bechah, Yassina et al. 2010. "Genomic, Proteomic, and Transcriptomic Analysis of Virulent and Avirulent *Rickettsia Prowazekii* Reveals Its Adaptive Mutation Capabilities.." *Genome research* 20(5): 655–63.

Bierne, Hélène, and Pascale Cossart. 2012. "When Bacteria Target the Nucleus: the Emerging Family of Nucleomodulins.." *Cellular Microbiology* 14(5): 622–33.

Brickman, Timothy J et al. 2011. "Transcriptional Profiling of the Iron Starvation Response in *Bordetella Pertussis* Provides New Insights Into Siderophore Utilization and Virulence Gene Expression.." *Journal of Bacteriology* 193(18): 4798–4812.

Carlson, Paul E et al. 2009. "Transcriptional Profiling of *Bacillus Anthracis* Sterne (34F2) During Iron Starvation.." *PLoS ONE* 4(9): e6988.

Chen, Chen et al. 2010. "Large-Scale Identification and Translocation of Type IV Secretion Substrates by *Coxiella Burnetii*.." *Proceedings of the National Academy of Sciences of the United States of America* 107(50): 21755–60.

Cheng, Chuanmin et al. 2013. "Targeted and Random Mutagenesis of *Ehrlichia Chaffeensis* For the Identification of Genes Required for *In Vivo* Infection.." *PLoS Pathogens* 9(2): e1003171.

Cheng, Zhihui, Mingqun Lin, and Yasuko Rikihisa. 2014. "*Ehrlichia Chaffeensis* Proliferation Begins with NtrY/NtrX and PutA/GlnA Upregulation and CtrA Degradation Induced by Proline and Glutamine Uptake.." *mBio* 5(6): e02141–14–e02141–14.

Cheng, Zhihui, Xueqi Wang, and Yasuko Rikihisa. 2008. "Regulation of Type IV Secretion Apparatus Genes During *Ehrlichia Chaffeensis* Intracellular Development by a Previously Unidentified Protein.." *Journal of Bacteriology* 190(6): 2096–2105.

Cossart, Pascale, and Alice Lebreton. 2014. "A Trip in the 'New Microbiology' with the Bacterial Pathogen *Listeria Monocytogenes*.." *FEBS letters* 588(15): 2437–45.

Cunnac, Sébastien et al. 2011. "Genetic Disassembly and Combinatorial Reassembly Identify a Minimal Functional Repertoire of Type III Effectors in *Pseudomonas Syringae*.." *Proceedings of the National Academy of Sciences of the United States of America* 108(7): 2975–80.

De Jong, Maarten F et al. 2008. "Identification of VceA and VceC, Two Members of the VjbR Regulon That Are Translocated Into Macrophages by the *Brucella* Type IV Secretion System.." *Molecular Microbiology* 70(6): 1378–96.

de Vries, Stefan P W et al. 2013. "Genetic Requirements for *Moraxella Catarrhalis* Growth Under Iron-Limiting Conditions.." *Molecular Microbiology* 87(1): 14–29.

Escolar, L, J Pérez-Martín, and V de Lorenzo. 1998. "Binding of the Fur (Ferric Uptake Regulator) Repressor of *Escherichia Coli* To Arrays of the GATAAT Sequence.." *Journal of molecular biology* 283(3): 537–47.

Felsheim, Roderick F et al. 2010. "Transformation of *Anaplasma Marginale*.." *Veterinary Parasitology* 167(2-4): 167–74.

Ferguson, Andrew D, and Johann Deisenhofer. 2004. "Metal Import Through Microbial Membranes.." *Cell* 116(1): 15–24.

Galdiero, Stefania et al. 2012. "Microbe-Host Interactions: Structure and Role of Gram-Negative Bacte-Rial Porins." *Current Protein and Peptide Science* 13(8): 843–54.

Garcia-Garcia, Jose C et al. 2009. "Silencing of Host Cell *Cybb* Gene Expression by the Nuclear Effector AnkA of the Intracellular Pathogen *Anaplasma Phagocytophilum*.." *Infection and immunity* 77(6): 2385–91.

Hajri, Ahmed et al. 2009. "A 'Repertoire for Repertoire' Hypothesis: Repertoires of Type Three Effectors Are Candidate Determinants of Host Specificity in *Xanthomonas*.." *PLoS ONE* 4(8): e6632.

Hakimi, Mohamed-ali, and Dominique Cannella. 2011. "Apicomplexan Parasites and Subversion of the Host Cell microRNA Pathway.." *Trends in Parasitology* 27(11): 481–86.

Huang, Haibin et al. 2007. "Porin Activity of *Anaplasma Phagocytophilum* Outer Membrane Fraction and Purified P44.." *Journal of Bacteriology* 189(5): 1998–2006.

Ijdo, J W et al. 1998. "Cloning of the Gene Encoding the 44-Kilodalton Antigen of the Agent of Human Granulocytic Ehrlichiosis and Characterization of the Humoral Response.." *Infection and immunity* 66(7): 3264–69.

IJdo, Jacob W, Adam C Carlson, and Elizabeth L Kennedy. 2007. "*Anaplasma Phagocytophilum* AnkA Is Tyrosine-Phosphorylated at EPIYA Motifs and Recruits SHP-1 During Early Infection.." *Cellular Microbiology* 9(5): 1284–96.

Ivanov, Stanimir S, and Craig R Roy. 2009. "Modulation of Ubiquitin Dynamics and Suppression of DALIS Formation by the *Legionella Pneumophila* Dot/Icm System.." *Cellular Microbiology* 11(2): 261–78.

Jakubowski, Simon J, Eric Cascales, Vidhya Krishnamoorthy, and Peter J Christie. 2005. "*Agrobacterium Tumefaciens* VirB9, an Outer-Membrane-Associated Component of a Type IV Secretion System, Regulates Substrate Selection and T-Pilus Biogenesis.." *Journal of Bacteriology* 187(10): 3486–95.

Jongejan, F et al. 1991. "The Tick-Borne Rickettsia Cowdria Ruminantium Has a Chlamydia-Like Developmental Cycle.." *The onderstepoort journal of veterinary research* 58(4): 227–37.

Kirzinger, Morgan W B, and John Stavrinides. 2012. “Host Specificity Determinants as a Genetic Continuum.” *Trends in Microbiology* 20(2): 88–93.

Knowles, Timothy J, Anthony Scott-Tucker, Michael Overduin, and Ian R Henderson. 2009. “Membrane Protein Architects: the Role of the BAM Complex in Outer Membrane Protein Assembly..” *Nature Publishing Group* 7(3): 206–14.

Koßmehl, Sebastian et al. 2013. “Subcellular Protein Localization (Cell Envelope) in *Phaeobacter Inhibens* Dsm 17395..” *PROTEOMICS* 13(18-19): 2743–60.

Krewulak, Karla D, and Hans J Vogel. 2011. “TonB or Not TonB: Is That the Question?.” *Biochemistry and cell biology = Biochimie et biologie cellulaire* 89(2): 87–97.

Lenhart, Tiffany R, and Darrin R Akins. 2010. “*Borrelia Burgdorferi* Locus BB0795 Encodes a BamA Orthologue Required for Growth and Efficient Localization of Outer Membrane Proteins..” *Molecular Microbiology* 75(3): 692–709.

Li, Zhiru, and Clotilde K S Carlow. 2012. “Characterization of Transcription Factors That Regulate the Type IV Secretion System and Riboflavin Biosynthesis in *Wolbachia* Of *Brugia Malayi*..” *PLoS ONE* 7(12): e51597.

Lin, Mingqun, and Yasuko Rikihisa. 2003. “*Ehrlichia Chaffeensis* And *Anaplasma Phagocytophilum* Lack Genes for Lipid a Biosynthesis and Incorporate Cholesterol for Their Survival..” *Infection and immunity* 71(9): 5324–31.

Liu, Hongyan et al. 2012. “*Ehrlichia* Type IV Secretion Effector ECH0825 Is Translocated to Mitochondria and Curbs ROS and Apoptosis by Upregulating Host MnSOD..” *Cellular Microbiology* 14(7): 1037–50.

Lockwood, Svetlana et al. 2011. “Identification of *Anaplasma Marginale* Type IV Secretion System Effector Proteins..” *PLoS ONE* 6(11): e27724.

Lopez, Job E et al. 2007. “Immunogenicity of *Anaplasma Marginale* Type IV Secretion System Proteins in a Protective Outer Membrane Vaccine..” *Infection and immunity* 75(5): 2333–42.

Marcelino, I et al. 2012. “Proteomic Analyses of *Ehrlichia Ruminantium* Highlight Differential Expression of MAP1-Family Proteins..” *Veterinary Microbiology* 156(3-4): 305–14.

Marchesini, María Inés et al. 2011. “In Search of *Brucella Abortus* Type IV Secretion Substrates: Screening and Identification of Four Proteins Translocated Into Host Cells Through VirB System..” *Cellular Microbiology* 13(8): 1261–74.

Martinez, Eric et al. 2014. “Identification of OmpA, a *Coxiella Burnetii* Protein Involved in Host Cell Invasion, by Multi-Phenotypic High-Content Screening..” *PLoS Pathogens* 10(3): e1004013.

Martínez-Núñez, Carola et al. 2010. “The Two-Component System BvrR/BvrS Regulates the Expression of the Type IV Secretion System VirB in *Brucella Abortus*..” *Journal of Bacteriology* 192(21): 5603–8.

Mellin, J R et al. 2014. “Riboswitches. Sequestration of a Two-Component Response Regulator by a Riboswitch-Regulated Noncoding RNA..” *Science* 345(6199): 940–43.

Meyer, Damien F et al. 2013. “Searching Algorithm for Type IV Secretion System Effectors 1.0: a Tool for Predicting Type IV Effectors and Exploring Their Genomic Context..” *Nucleic Acids Research* 41(20): 9218–29.

Miyata, Sarah T, Verena Bachmann, and Stefan Pukatzki. 2013. “Type VI Secretion System Regulation as a Consequence of Evolutionary Pressure..” *Journal of medical microbiology* 62(Pt 5): 663–76.

Mohan Kumar, Dipu et al. 2013. “*Ehrlichia Chaffeensis* Uses Its Surface Protein EtpE to Bind GPI-Anchored Protein DNase X and Trigger Entry Into Mammalian Cells..” *PLoS Pathogens* 9(10): e1003666.

Navarro, Lionel et al. 2008. “Suppression of the microRNA Pathway by Bacterial Effector Proteins..” *Science*

321(5891): 964–67.

Nelson, Curtis M et al. 2008. “Whole Genome Transcription Profiling of *Anaplasma Phagocytophilum* In Human and Tick Host Cells by Tiling Array Analysis..” *bmc genomics* 9(1): 364.

Newton, Hayley J, Justin A McDonough, and Craig R Roy. 2013. “Effector Protein Translocation by the *Coxiella Burnetii* Dot/Icm Type IV Secretion System Requires Endocytic Maturation of the Pathogen-Occupied Vacuole..” *PLoS ONE* 8(1): e54566.

Niu, Hua et al. 2012. “Autophagosomes Induced by a Bacterial Beclin 1 Binding Protein Facilitate Obligatory Intracellular Infection..” *Proceedings of the National Academy of Sciences of the United States of America* 109(51): 20800–20807.

Niu, Hua, Vera Kozjak-Pavlovic, Thomas Rudel, and Yasuko Rikihisa. 2010. “*Anaplasma Phagocytophilum* Ats-1 Is Imported Into Host Cell Mitochondria and Interferes with Apoptosis Induction..” *PLoS Pathogens* 6(2): e1000774.

Noh, Susan M et al. 2008. “Composition of the Surface Proteome of *Anaplasma Marginale* And Its Role in Protective Immunity Induced by Outer Membrane Immunization..” *Infection and immunity* 76(5): 2219–26.

Nyika, A et al. 1998. “A DNA Vaccine Protects Mice Against the Rickettsial Agent *Cowdria Ruminantium*..” *Parasite immunology* 20(3): 111–19.

Ohashi, N, N Zhi, Y Zhang, and Y Rikihisa. 1998. “Immunodominant Major Outer Membrane Proteins of *Ehrlichia Chaffeensis* Are Encoded by a Polymorphic Multigene Family..” *Infection and immunity* 66(1): 132–39.

Oren, Yaara et al. 2014. “Transfer of Noncoding DNA Drives Regulatory Rewiring in Bacteria..” *Proceedings of the National Academy of Sciences of the United States of America* 111(45): 16112–17.

Porte, F, J P Liautard, and S Köhler. 1999. “Early Acidification of Phagosomes Containing *Brucella Suis* Is Essential for Intracellular Survival in Murine Macrophages..” *Infection and immunity* 67(8): 4041–47.

Portier, Emilie et al. 2014. “*IroT*/*mavN*, A New Iron-Regulated Gene Involved in *Legionella Pneumophila* Virulence Against Amoebae and Macrophages..” *Environmental Microbiology*.

Postigo, M et al. 2007. “Differential Transcription of the Major Antigenic Protein 1 Multigene Family of *Ehrlichia Ruminantium* In *Amblyomma Variegatum* Ticks..” *Veterinary Microbiology* 122(3-4): 298–305.

Pretorius, A et al. 2010. “Studies of a Polymorphic *Ehrlichia Ruminantium* Gene for Use as a Component of a Recombinant Vaccine Against Heartwater.” *Vaccine* 28(20): 3531–39.

Pruneau, Ludovic et al. 2012. “Global Gene Expression Profiling of *Ehrlichia Ruminantium* At Different Stages of Development..” *FEMS Immunology & Medical Microbiology* 64(1): 66–73.

Pruneau, Ludovic et al. 2014. “Understanding *Anaplasmataceae* Pathogenesis Using ‘Omics’ Approaches..” *Frontiers in cellular and infection microbiology* 4(86): 1–7.

Rikihisa, Yasuko. 2010. “*Anaplasma Phagocytophilum* And *Ehrlichia Chaffeensis*: Subversive Manipulators of Host Cells.” *Nature Publishing Group* 8(5): 328–39.

Rikihisa, Yasuko, and Mingqun Lin. 2010. “*Anaplasma Phagocytophilum* And *Ehrlichia Chaffeensis* Type IV Secretion and Ank Proteins..” *Current Opinion in Microbiology* 13(1): 59–66.

Romano, Patricia S et al. 2007. “The Autophagic Pathway Is Actively Modulated by Phase II *Coxiella Burnetii* To Efficiently Replicate in the Host Cell..” *Cellular Microbiology* 9(4): 891–909.

Rüter, Christian, and Philip R Hardwidge. 2013. “‘Drugs From Bugs’: Bacterial Effector Proteins as Promising Biological (Immune-) Therapeutics.” *FEMS Microbiology Letters* 351(2): 126–32.

Schulein, Ralf et al. 2005. “A Bipartite Signal Mediates the Transfer of Type IV Secretion Substrates of *Bartonella*

Henselae Into Human Cells.." *Proceedings of the National Academy of Sciences of the United States of America* 102(3): 856–61.

Sieira, Rodrigo et al. 2012. "A MarR-Type Regulator Directly Activates Transcription From the *Brucella Abortus virB* Promoter by Sharing a Redundant Role with HutC.." *Journal of Bacteriology* 194(23): 6431–40.

Skaar, Eric P. 2010. "The Battle for Iron Between Bacterial Pathogens and Their Vertebrate Hosts.." *PLoS Pathogens* 6(8): e1000949.

Steele-Mortimer, Olivia. 2008. "The *Salmonella*-Containing Vacuole: Moving with the Times.." *Current Opinion in Microbiology* 11(1): 38–45.

Sutten, Eric L et al. 2010. "*Anaplasma Marginale* Type IV Secretion System Proteins VirB2, VirB7, VirB11, and VirD4 Are Immunogenic Components of a Protective Bacterial Membrane Vaccine.." *Infection and immunity* 78(3): 1314–25.

Teixidó, Laura et al. 2011. "Fur Activates the Expression of Salmonella Enterica Pathogenicity Island 1 by Directly Interacting with the *hilD* Operator *In Vivo* And *In Vitro*.." *PLoS ONE* 6(5): e19711.

Tseng, Tsai-Tien, Brett M Tyler, and João C Setubal. 2009. "Protein Secretion Systems in Bacterial-Host Associations, and Their Description in the Gene Ontology.." *BMC Microbiology* 9 Suppl 1(Suppl 1): S2.

van Kleef, M, A W Neitz, and D T De Waal. 1993. "Isolation and Characterization of Antigenic Proteins of *Cowdria Ruminantium*.." *Revue d'élevage et de médecine vétérinaire des pays tropicaux* 46(1-2): 157–64.

van Vliet, A H, F Jongejan, M van Kleef, and B A van der Zeijst. 1994. "Molecular Cloning, Sequence Analysis, and Expression of the Gene Encoding the Immunodominant 32-Kilodalton Protein of *Cowdria Ruminantium*.." *Infection and immunity* 62(4): 1451–56.

Vergunst, Annette C et al. 2005. "Positive Charge Is an Important Feature of the C-Terminal Transport Signal of the VirB/D4-Translocated Proteins of *Agrobacterium*.." *Proceedings of the National Academy of Sciences of the United States of America* 102(3): 832–37.

Viadas, Cristina et al. 2010. "Transcriptome Analysis of the *Brucella Abortus* BvrR/BvrS Two-Component Regulatory System.." *PLoS ONE* 5(4): e10216.

Voth, Daniel E, Laura J Broederdorf, and Joseph G Graham. 2012. "Bacterial Type IV Secretion Systems: Versatile Virulence Machines.." *Future Microbiology* 7(2): 241–57.

Wang, Xin et al. 2009. "Edwardsiella Tarda T6SS Component *evpP* Is Regulated by *esrB* And Iron, and Plays Essential Roles in the Invasion of Fish.." *Fish & shellfish immunology* 27(3): 469–77.

Wang, Xueqi et al. 2007. "*Anaplasma Phagocytophilum* P44 mRNA Expression Is Differentially Regulated in Mammalian and Tick Host Cells: Involvement of the DNA Binding Protein ApxR.." *Journal of Bacteriology* 189(23): 8651–59.

Wang, Xueqi, Takane Kikuchi, and Yasuko Rikihisa. 2007. "Proteomic Identification of a Novel *Anaplasma Phagocytophilum* DNA Binding Protein That Regulates a Putative Transcription Factor.." *Journal of Bacteriology* 189(13): 4880–86.

Wu, Chih-Feng, Jer-Sheng Lin, Gwo-Chyuan Shaw, and Erh-Min Lai. 2012. "Acid-Induced Type VI Secretion System Is Regulated by ExoR-ChvG/ChvI Signaling Cascade in *Agrobacterium Tumefaciens*.." *PLoS Pathogens* 8(9): e1002938.

Zhi, Ning et al. 2002. "Transcript Heterogeneity of the *P44* Multigene Family in a Human Granulocytic Ehrlichiosis Agent Transmitted by Ticks.." *Infection and immunity* 70(3): 1175–84.

Zhu, Wenhan et al. 2011. "Comprehensive Identification of Protein Substrates of the Dot/Icm Type IV Transporter of *Legionella Pneumophila*.." *PLoS ONE* 6(3): e17638.

ANNEXES

Nucleic Acids Research Advance Access published August 13, 2013

Nucleic Acids Research, 2013, 1–12
doi:10.1093/nar/gkt718

Searching algorithm for type IV secretion system effectors 1.0: a tool for predicting type IV effectors and exploring their genomic context

Damien F. Meyer[1,2,*], Christophe Noroy[1,2], Amal Moumène[1,2,3], Sylvain Raffaele[4,5], Emmanuel Albina[1,2] and Nathalie Vachiéry[1,2]

[1]CIRAD, UMR CMAEE, F-97170 Petit-Bourg, Guadeloupe, France, [2]INRA, UMR1309 CMAEE, F-34398, Montpellier, France, [3]Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre cedex, Guadeloupe, France, [4]INRA, Laboratoire des Interactions Plantes-Microorganismes, UMR441, Castanet-Tolosan, France and [5]CNRS, Laboratoire des Interactions Plantes-Microorganismes, UMR2594, Castanet-Tolosan, France

Received May 9, 2013; Revised July 19, 2013; Accepted July 22, 2013

ABSTRACT

Type IV effectors (T4Es) are proteins produced by pathogenic bacteria to manipulate host cell gene expression and processes, divert the cell machinery for their own profit and circumvent the immune responses. T4Es have been characterized for some bacteria but many remain to be discovered. To help biologists identify putative T4Es from the complete genome of α- and γ-proteobacteria, we developed a Perl-based command line bioinformatics tool called S4TE (searching algorithm for type-IV secretion system effectors). The tool predicts and ranks T4E candidates by using a combination of 13 sequence characteristics, including homology to known effectors, homology to eukaryotic domains, presence of subcellular localization signals or secretion signals, etc. S4TE software is modular, and specific motif searches are run independently before ultimate combination of the outputs to generate a score and sort the strongest T4Es candidates. The user keeps the possibility to adjust various searching parameters such as the weight of each module, the selection threshold or the input databases. The algorithm also provides a GC% and local gene density analysis, which strengthen the selection of T4E candidates. S4TE is a unique predicting tool for T4Es, finding its utility upstream from experimental biology.

INTRODUCTION

Bacterial pathogens have evolved specific effector proteins to exploit host cell machinery and hijack the immune responses during infection (1). Dedicated multiprotein complexes, known as secretion systems, secrete these effectors. Type IV secretion systems (T4SS) are specialized ATP-dependent protein complexes used by many bacterial pathogens for the delivery of type IV effector (T4E) proteins into eukaryotic cells to subvert host cell processes during infection. Some T4Es have been identified in α-proteobacteria (*Agrobacterium tumefaciens, Bartonella henselae*, *Brucella abortus*, *Anaplasma* spp. and *Ehrlichia chaffeensis*) and γ-proteobacteria (*Coxiella burnetii* and *Legionella pneumophila*) and shown to be critical for pathogenicity making them first choice targets to understand bacterial virulence (1–12). Our group was initially interested in identifying T4Es in *Ehrlichia ruminantium,* which is the causative agent of heartwater, a fatal tropical disease of ruminants. This α-proteobacterium belong to the Anaplasmataceae family and is transmitted by ticks of genus *Amblyomma* (13).

Ehrlichia spp. and *Anaplasma* spp. of the Anaplasmataceae family are obligate intracellular pathogens of humans and animals capable of infecting various cell types, including endothelial cells, granulocytes, monocytes and macrophages (14). Once inside the host cell, *Ehrlichia* spp and *Anaplasma* spp. reside inside a membrane-bound vacuole where they replicate (14). The replicative vacuole interacts with cholesterol and autophagosome pathways for maturation (15,16). The biogenesis of this replicative niche depends on the function of T4SS and the related secretion of T4Es (16). However, only two T4Es have been described so far in Anaplasmataceae family and shown to play an important role in invasion and pathogenesis. The first effector, AnkA, was identified in *Anaplasma phagocytophilum*, based on sequence homology with tandemly repeated ankyrin motifs (17). AnkA is secreted by T4SS, is tyrosine phosphorylated and is then addressed into the

*To whom correspondence should be addressed. Tel: +590 590 25 59 47; Fax: +590 590 94 03 96; Email: damien.meyer@cirad.fr

nucleus to silence the *CYBB* gene expression of the host cell (18–20). This effector is part of the emerging family of the nucleomodulins that hijack nuclear processes to facilitate infection (21). The other known Anaplasmataceae effector, Ats-1, was identified in *A. phagocytophilum* and shown to be targeted by T4SS to the cytoplasm of infected cells. Ats-1 interacts with the host autophagosome initiation complex to recruit autophagosomes to the bacterial intracellular vacuole (16). Another portion of Ats-1 targets host cell mitochondria to exert antiapoptotic activity (12,22)

To facilitate the identification of putative T4Es in the whole genome of *E. ruminantium*, we explored bioinformatics for sequence motif search. Bacterial effectors can be classified into two main groups: those mimicking endogenous cell proteins and those modifying host cell proteins (1,23,24). Because of their various functions in mammalian cells, most of these effectors have characteristic eukaryotic-like domains involved in protein–protein interactions, localization signals and C-terminal features like positive charge, basicity or hydrophobicity that interfere with host cellular processes to promote bacterial replication (12,14,16,25). We first looked at the literature for bioinformatics tools developed for such motif searches and tested successfully for T4Es prediction (4,6,9–11, 25–28). However, all the algorithms used in these works were based on the identification of a limited number of sequence motifs and were designed for specific α- or γ-proteobacteria. To circumvent the risk of missing important T4Es, we decided to develop a new algorithm combining searches for all motifs that previously predicted T4Es of α- and γ-proteobacteria. In addition, new sequence motifs were derived from the analysis of the extensive repertoire of T4Es characterized in *L. pneumophila* (8) and included in the algorithm.

In this article, we present 'S4TE' (Searching Algorithm for Type-IV secretion system Effectors), a tool for *in silico* screening of proteobacteria genomes and T4Es prediction based on the combined use of 13 distinctive features. This software was first probed against the comprehensive T4E dataset of *L. pneumophila,* strain Philadelphia (8) and subsequently tested on several genomes of α- and γ-proteobacteria. S4TE is both memory- and time-efficient. Although advanced users will be capable of modifying searching parameters of S4TE (e.g. exclusion of modules, change in module weighting, selection threshold or input databases), the common user can easily run the program with default settings. Installation process and basic command lines to launch and run S4TE are detailed in the user guide. S4TE package is freely available to non-commercial users at http://sate.cirad.fr/.

MATERIALS AND METHODS

Overview

We propose an easy-to-use and customizable algorithm for the prediction of candidate effector proteins secreted by T4SS. The algorithm can be used as a standard pre-selection technique for T4 effectors in genomes of any size. Its modularity will offer a simple and robust alternative to machine learning approach for less-studied pathogenic bacteria. In this section, we describe the algorithm used by S4TE, how the parameters of this software were estimated from the literature and how S4TE performs on different genomes. The essential features of the S4TE program, as depicted in Figure 1, are the following: (i) genome-wide screening based on 13 different criteria including homology to known T4Es, occurrence of eukaryotic-like domains or motifs and subcellular localization signals; (ii) T4Es prediction and ordering output based on criteria scoring; (iii) information on prediction performance compared with the reference *L. pneumophila*; (iv) analysis of G+C content and of space clustering of S4TE hits and (v) analysis of genome architecture and distribution of S4TE hits depending on local gene density.

S4TE search modules

S4TE is a modular program written in Perl that screens bacterial genomes. The 13 search parameters are listed in Table 1 and are run in 10 modules detailed below.

(1) *De novo regulatory motif search*. In *Legionella*, the PmrA and CpxR response regulators regulate numerous T4Es (40). Because *cis*-acting regulatory sequences have not been described so far in α-proteobacteria, a motif search was performed in the promoter region of the 19 known T4Es of this class of bacteria in *Anaplasma*, *Ehrlichia*, *Brucella* and *Bartonella* genera. Enriched DNA motifs were searched in a window of 300 nt placed upstream of the start codon, using MEME (41) (http://meme.nbcr.net/meme/). A consensus motif of 10 nt was identified in 14 promoters. This motif, termed RS-TY, consists of 3 purines (R), 1 strong base G or C (S), any nucleotide (A, T, G, C), 4 thymines (T) and 1 pyrimidine (Y) (Supplementary Figure S1). Interestingly, this motif is reminiscent of the *cis*-regulatory elements characterized in *L. pneumophila* that are required for expression of T4SS-encoding genes (42). Also, for other pathogenic bacteria, the expression of genes encoding secretion systems and those scattered in the genome encoding their substrates is co-regulated by one master regulatory protein that binds a consensus or imperfect *cis*-regulatory element (43). Although the biological significance (if any) of this motif needs to be tested experimentally, it was included in S4TE algorithm. The corresponding *RS_TY.pl* module will extract the 300 nt upstream from the START codon and searches for the RS-TY motif thanks to a position-specific scoring matrix generated from multiple sequence alignments with the promoters of known T4Es of α-proteobacteria. Only alignments with a score >130 are selected.

(2) *Homology*. Effectors were shown not only to share local sequence similarities but also to diverge rapidly (8). BLAST 2.2 (44) was used for protein comparisons to look for homologies with known T4Es. The command line application can be downloaded from the NCBI web page. The blastp software allows the search of local similarity regions between two protein sequences. S4TE compares the database containing all known T4Es with the query proteome and returns all proteins containing a region that has local similarity with a cutoff of expected

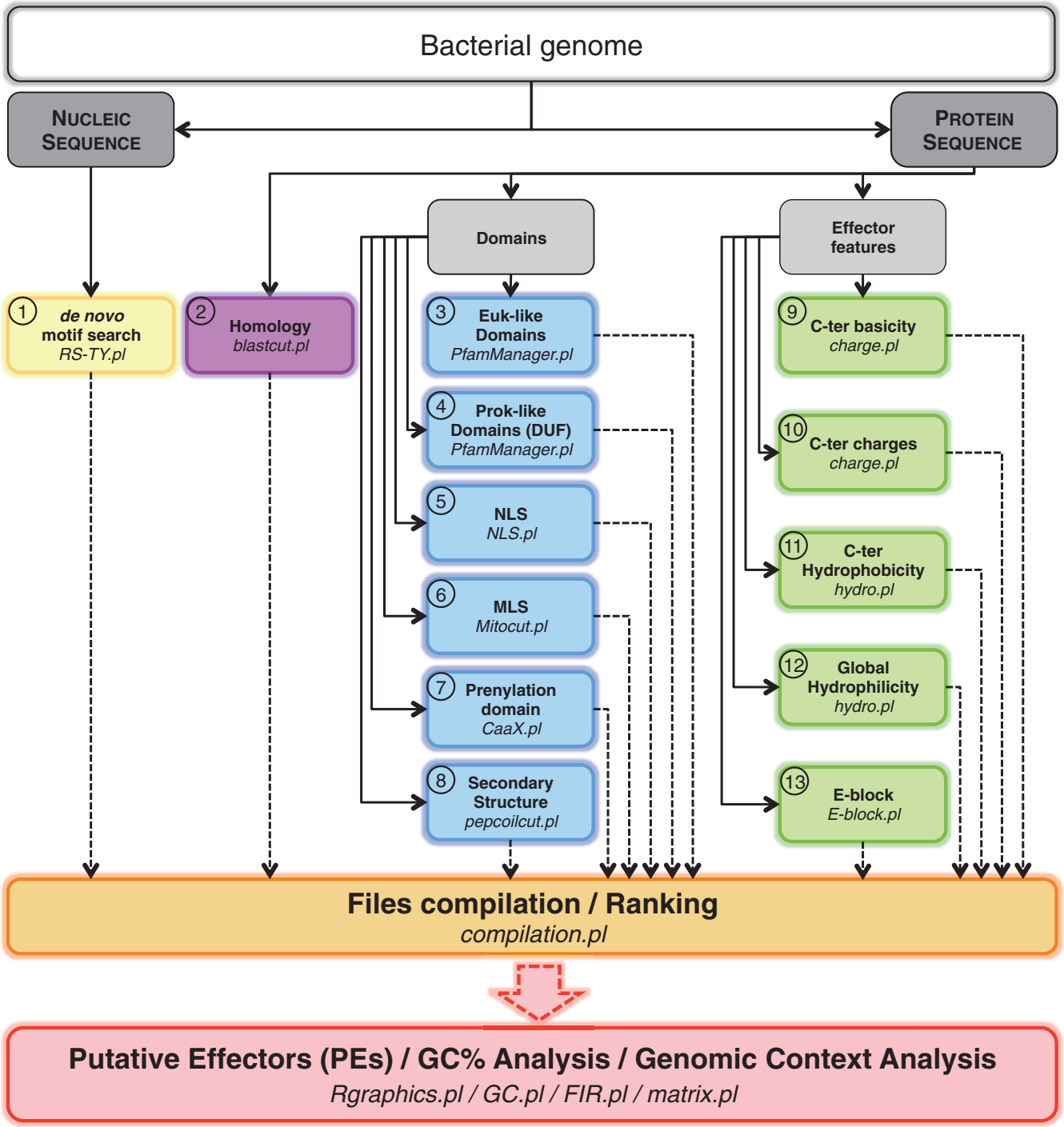


Figure 1. Flowchart of the bioinformatics search by S4TE to identify putative effector proteins (PEs). This bioinformatics pipeline is composed of 15 steps delimited by color boxes. Steps 1–13 look for T4 effector features in a given bacterial genome. Step 14 (Files compilation/Ranking) ranks and classifies the predicted T4 effectors based on their number of features to provide the best candidates for experimental validation (Supplementary Figures S2 and S3). Step 15 analyses the genome architecture and G + C content and shows the distribution of predicted effectors. Programs used are indicated in italics. Euk-like, Eukaryotic-like; Prok-like, Prokaryotic-like; NLS, nuclear localization signal; MLS, mitochondrial localization signal; C-ter, C-terminal.

value (E) <0.01. This E-value cutoff of 0.01 was selected to show similarities between phylogenetically distant bacterial species (6). The majority of known T4Es are from γ-proteobacteria, and S4TE was expected to work also with α-proteobacteria. The *blastcut.pl* program was written to reformat the output file of blastp program for user-friendly reading and for S4TE compiler. The *blastcut.pl* returns only blastp positive alignments.

(3) and (4) *Eukaryotic- and prokaryotic-like domains.* The search for eukaryotic- and prokaryotic-like domains

Table 1. Description of the 13 features used in S4TE to screen a bacterial genome

Feature number	Feature name	Description	References
1	*De novo* motif search	RRRSNTTTTY motif in the −300 bp (Supplementary Figure S1)	This work, http://meme.sdsc.edu/meme/cgi-bin/meme.cgi
2	Homology	Sequence identity to a known effector molecule; Blastp against effector database (e-value = 10^{-2})	(6,8)
3	Euk-like domains	Presence of eukaryotic domain: 58 eukaryotic domains	This work, Supplementary Table S1
4	Prok-like domains	3617 Domain of Unknown Function (DUF) domains	(29), http://pfam.sanger.ac.uk/search/
5	NLS (nuclear localization signal)	Monopartite NLS; [KR]-[KR]-[KR]-[KR]-[KR] and bipartite NLS; K-[KR]-X(6,20)-[KR]-[KR]-X-[KR]	(30,31)
6	MLS (mitochondrial localization signal)	Probability of a sequence containing a mitochondrial targeting peptide ($P > 0.95$)	(32), http://www.bioperl.org/
7	Prenylation domain	CaaX at the C-terminal; 'C' represents a cysteine residue, 'a' denotes an aliphatic amino acid and 'X' is one of four amino acids	(1,33,34)
8	Secondary structure Coiled coils	Probability of a coiled-coil structure for windows of 28 residues through a protein sequence ($P > 0.95$)	(35,36), http://emboss.bioinformatics.nl/cgi-bin/emboss/pepcoil
9	C-ter basicity	≤3 [HRK] in the C-terminal 25 amino acids	(3,7,22)
10	C-ter charges	Charge of C-terminal 25 amino acids ≥2; C-ter charge = number of [HRK]-number of [ED]-1 (COO^-)	(3,7,22)
11	C-ter hydrophobicity	Hydropathy of C-terminal 25 amino acids; Hydrophobic residue at the −3rd or -4^{th} position	(9,11,37,38)
12	Global hydrophilicity	Hydropathy of total protein <−200	(9,11,37)
13	E-block	EEXXE in the C-terminal 30 amino acids	(39)

is done by *Pfam-scan.pl* in the PfamScan package (http://pfam.sanger.ac.uk) (45). To run properly, *Pfam-scan.pl* needs several softwares: the Moose module in CPAN (http://search.cpan.org/~ether/Moose-2.0801/lib/Moose.pm), hmmer3.0rc2 and the BioPerl-1.6.1.tar.gz package. In addition, *Pfam-scan.pl* needs motif database Pfam-A.hmm (http://pfam.sanger.ac.uk). For file size purpose and memory saving, *hmmcut.pl* was designed to generate a motif database with the Pfam ID of each motif of interest. This motif database contains 58 eukaryotic-like domains previously found in effectors (Supplementary Table S1) and 3617 prokaryotic-like DUF domains (29). *PfamManager.pl* will reformat the output file of Pfam program for user convenience by separating search results for eukaryotic-like domains and for DUF domains.

(5) *Nuclear localization signals (NLS)*. NLS are protein sequences that target proteins in the nucleus of eukaryotic cells (1). We assumed that the occurrence of NLS in a bacterial protein sequence would be a good indicator of secretion. There are two classes of NLS. Monopartite NLS consist of the PKKKRKV motif (30). Bipartite NLS are more complex and consist of two alkaline clusters (K and R) separated by a variable spacer (31). We wrote *NLS.pl* to search for monopartite NLS with the [KR]-[KR]-[KR]-[KR]-[KR] motif and for bipartite NLS with the K-[KR]-X(6,20)-[KR]-[KR]-X-[KR] motif. The latter motif was derived from multiple alignments of known eukaryotic protein sequences containing NLS (31).

(6) *Mitochondrial localization signals (MLS)*. MLS are signal sequences located in the N-terminus of proteins that are targeted to mitochondria. This sequence is cleaved after translocation of the protein inside the mitochondria (1,22). To predict MLS and extract the predicted signal, we used the *Mitoprot.pm* package of Bioperl (32). *Mitoprot.pl* and *Mitocut.pl* were developed to use the Perl module *Mitoprot.pm* and to format the output file, respectively. Only MLS with $P > 0.95$ are selected by S4TE.

(7) *Prenylation domains*. Prenylation is a permanent post-translational modification that is required for protein stability (1). Prenylation involves the covalent addition of a 15-carbon farnesyl or a 20-carbon geranyl-geranyl isoprenoid group to a Cys residue within the conserved C-terminal CaaX motif (in which 'a' represents an aliphatic residue and 'X' is one of the four amino acids) (33). Prenylation increases protein hydrophobicity, facilitating protein anchorage to membranes and targeting effector proteins to membrane-bound organelles (34). S4TE module *CaaX.pl* will search for prenylation domain in the C-terminal of proteins.

(8) *Coiled coils*. Coiled coils are structural motifs in proteins in which at least two α-helices are coiled together (35). Coiled-coil domains are protein interaction domains and have a role in the regulation of gene expression by stabilizing transcription factors (36). Coiled-coil–type proteins have similarities with pore-forming proteins in gram-negative pathogens and seem to be important for the delivery of effectors into host cells (46). Most proteins containing validated coiled-coil domains are of eukaryotic origin (47). Interestingly, coiled-coil domains are frequently found in secreted virulence effector proteins (47). To search for coiled-coil domains, we used *pepcoil* software of Emboss package (http://emboss.bioinformatics.nl/cgi-bin/emboss/pepcoil). The module *pepcoilcut.pl* of S4TE extracts coiled-coil domains with $P > 0.95$ and formats the output file.

(9) and (10) *C-terminal basicity and charge*. T4Es often have a positive charge and a large number of alkaline amino acids (HRK) in the 25 C-terminal amino acids (3,7,22).

In S4TE, these two features are investigated with *charge.pl* module. In α-proteobacteria, all known T4Es have three or more alkaline amino acids (HRK) in the 25 C-terminal amino acids. This feature was used as a threshold to select positives. Charge is calculated by summing the positively charged amino acids (HRK) and by subtracting the number of negatively charged amino acids (ED) and the negative C-terminal charge (COO^-). In α-proteobacteria, most known T4Es have a C-terminal charge of at least 2. This value was set as a threshold in *charge.pl*.

(11) and (12) *C-terminal and global hydropathy*. We calculated hydropathy profiles of proteins using Kyte–Doolittle scale, for which the more hydrophobic the residue, the higher its hydropathy value (>0) (37). Most known T4Es are hydrophilic, having total negative hydropathy scores, negative average hydropathy and highly hydrophilic C-termini (9,11,38). Negative hydropathy at the C-terminus and negative global hydropathy with scores of <200 were used in S4TE with *hydro.pl* for screening in bacterial proteomes. Moreover, *hydro.pl* looks for a hydrophobic residue at the third or fourth C-terminal positions (9,48).

(13) *E-block*. The E-block domain consists of a glutamate-rich sequence (EEXXE) in the C-terminal 30 amino acids and is associated with T4Es translocation in *L. pneumophila*. Huang *et al.* showed that an E-block motif is also important for the translocation of T4SS substrates (39). This motif is searched by S4TE with the *E-block.pl* module.

All search modules return a score of 0 or 1, based on the absence or presence of the parameter, or a parameter quantitative value over the threshold. The individual scores, weighted or not according to user decision, are summed, and the global score is compared with a threshold set by the user. All hits higher than the threshold are returned by S4TE. Each search module was individually designed and refined with dedicated T4Es datasets of different origins to achieve the best specificity (i.e. probability to find a true negative [TN]). We then adjusted S4TE (module weights and threshold) to have the best positive predictive value (PPV; proportion of true positives [TP] detected among all positives) and sensitivity (probability to find a TP) on the dataset of 275 T4Es characterized in *L. pneumophila*, strain Philadelphia (see Results section). S4TE is also designed to provide two performance indicators attached to any combination of searched parameters. The first indicator is the Sensitivity Index for *Legionella* (SI_L) of a given combination: SI_L is calculated by $SI_L = \frac{TP}{(TP+FP)}$ and ranges from 0 to 1. SI_L gives information on the proportion of TP found in *L. pneumophila* with the same combination of parameters. The second indicator is defined as a Positivity Index for *Legionella* (PI_L) and is generated by the formula $PI_L = \frac{(TP-FP)}{FP}$. Therefore, this indicator integrates the net yield of TP given by a combination of parameters, in proportion to the FP number. PI_L ranges from −1 to 10 (Supplementary Table S2). When the FP number equals 0, PI_L indicator is 0. By using S4TE on different genomes, the user will be able to look at the *L. pneumophila* performances of all the combinations that picked T4Es hits. For instance, if four hits are selected with combinations that give the following values for {TP, FP} in *L. pneumophila*; {25, 9}, {6, 1}, {0, 2} and {0, 0}; the corresponding SI_L and PI_L will be {0.74, 1.78}; {0.86, 5}; {0, −2} and {0, 0}, respectively. In this situation, to select the first target for further biological validation, the hit with $SI_L = 0.86$ and $PI_L = 5$ could be considered as more promising than the one with $SI_L = 0.74$ and $PI_L = 1.78$ (increased risk of FP selection with the latter). In contrast, a hit obtained with a combination that does not exist in *L. pneumophila* ($SI_L = 0$ and $PI_L = 0$) could be interesting to test because it could serve to identify an effector with an original association of features.

Data compilation

The main program *S4TE.pl* executes the 10 previously described modules to generate 13 result files. Afterward, a compiler (*compilation.pl*) will collect all information from result files generated during the pipeline execution and will highlight important data (Figure 1). For each feature, *compilation.pl* will search the identified proteins and will count the number of positive hits. Then, *compilation.pl* will sort the results by top/down scoring. For each analysis, the compiled results are written in *CompilationFile.txt* with the list of identified proteins, the number of hits, the score and the combination of positive features for each protein.

S4TE graphical outputs

G + C content and space clustering analysis

In *L. pneumophila*, T4Es have atypical G + C content that could result from horizontal gene transfers (HGT) acquired during evolution (26). In addition, effectors encoding genes are often clustered in specific regions, indicating possible HGT events and suggesting possible co-regulation (27,49). With the *GC.pl* program, S4TE calculates the G + C content (GC%) in a 10-kb window sliding every 200 nt across the genome and plots the GC%, the mean GC% and candidate effectors: effectors in regions with high G + C content are plotted in red, whereas others are plotted in green (Figure 3). This representation allows the user to easily see whether the hits are clustered or scattered in the genome and whether they are rich or low in G + C content.

Genome architecture analysis

S4TE also proposes to analyse the genome architecture and its hit content through the visualization of the length and distribution of intergenic regions and the distribution of the hits according to local gene density (50–52). For every gene, we used two-dimensional data binning to visualize the distance to its closest coding gene neighbors on five prime and three prime (designated as 5′ and 3′ flanking intergenic regions [FIRs]) in a single representation (50). With *FIR.pl* and *matrix.pl*, S4TE sorts genes (or predicted effectors) into two-dimensional bins defined by the length of their 5′ and 3′ FIRs. Then, the gene (or effector) density distribution is represented in R by a color-coded heat map with *filled.contour.pl*. We used

the median length of FIRs to distinguish between gene-dense regions (GDRs) and gene-sparse regions (GSRs). This method offers the opportunity to visualize the position of predicted effectors relative to the whole genome architecture (52).

Databases

To analyse a given genome, S4TE needs a genome database with four distinct files: (i) `Genome.nucl` containing the FASTA genome sequence; (ii) `Genome.an`, a csv file (with; as separator) containing the gene ID; the position of the first nucleotide of coding sequence; the position of the last nucleotide; the sense or antisense status; before use in S4TE, this file needs to be sorted in ascending order from the first nucleotide position; (iii) `Genome.prot`, a fasta file containing all the protein sequences of the genome; and (iv) `Genome.csv`, a file constructed from `Genome.prot` with *nomprot.pl* in the folder `~/S4TE/DataBases/Genome/Tools/`

A database of validated effectors derived from the literature was constructed in a protein sequence file of effectors (*effector_db.txt*) and formatted by `makeblastdb -in effector_db.txt -dbtype prot`

The database assembling the eukaryotic Pfam domains found in T4Es was constructed from the downloaded original `wget ftp://ftp.sanger.ac.uk/pub/databases/Pfam/releases/Pfam26.0/Pfam-A.hmm*.gz`

The tool *hmmcut.pl* will construct the Pfam-A.hmm database with the motifs used by S4TE only. The Pfam-A.hmm database is formatted by `.hmmpress Pfam-A.hmm`.

Software availability

S4TE is freely available to non-commercial users at http://sate.cirad.fr/. Programming was done using Perl 5.12 and BioPerl 1.6. The software runs on Linux platforms (Ubuntu 11.10 and Mac OS X). All required packages and the installation process are described in the user guide included in Supplementary Methods S1. The user guide also details S4TE options for running S4TE. By default, the command line to launch S4TE is `.S4TE.pl -g ''name of the genome''` from the S4TE folder (`cd way_to_S4TE/S4TE/`). Some options are available for the user to launch S4TE: `-c`, suppression of a module in the pipeline; `-w`, modification of the weight of each module in the pipeline; `-t`, imposition of a threshold for effector selection. Each S4TE module creates an .txt file in the folder `way_to_S4TE/S4TE/Jobs/job<Name_of_genome_folder><year><month><day><hour><min>`

All the results are compiled in *CompilationFile.txt* and *Results.txt* in the same folder.

RESULTS

After the initial configuration of the different modules of the S4TE algorithm, achieved on all known T4Es, the whole program was run against the 275 known *L. pneumophila* T4Es and different parameter weightings were tested for optimizing the prediction. The next section details this optimization. The last section shows the outcome of S4TE on different representative bacteria genomes.

Effectors prediction and validation of the S4TE algorithm

For S4TE adjustment and validation, we used the extensive repertoire of 275 experimentally confirmed T4Es of *L. pneumophila*, strain Philadelphia (8,9). The work was carried out in three major steps consisting of the adjustment of parameter weighting for optimized prediction of *L. pneumophila* T4Es, the analysis of the relative importance of the different parameters in the optimized configuration and the link between the GC% and gene density in the genome and T4Es localization.

Optimization of S4TE forT4Es prediction on *L. pneumophila*

The genome of *L. pneumophila*, strain Philadelphia, contains the most extensive repertoire of T4Es ever identified, with 275 confirmed effector proteins encoded by 9.3% of the genome (8). These 275 T4Es were considered as TP, whereas the 2666 other proteins were included in the analysis as TN. By default, S4TE sets the selection threshold cutoff at 5. This threshold cutoff was first defined as the minimal score minus 1, necessary to allow the selection of all known T4Es of two α-proteobacteria: *A. tumefaciens* and *B. henselae*. The default threshold was selected to offset the overrepresentation of *L. pneumophila* effectors and to be less stringent for candidate effectors with no homology to known effectors. However, the user can adjust the threshold and S4TE will return all candidates scoring over the new threshold. On the *L. pneumophila* dataset, the algorithm with unweighted parameters and the default threshold at 5 selected 151 TP and 2428 TN. However, 238 hits were not T4Es (false positive [FP]), and 124 T4Es were not found by S4TE (false negative [FN]). This led to a sensitivity (Se) of 55%, a specificity (Sp) of 91%, PPV [PPV = TP/(TP + FP)] of 38.8% and a negative predictive value [NPV = TN/(TN + FN)] of 95.1%. To improve Se and PPV, different parameter weightings were tested. The best combination (i.e. 1311111111111) led to the selection of 223 TP of the 275 effectors of *L. pneumophila* (Se = 81%, PPV = 72.8%); it led to only 83 FP (Sp = 98.8%, NPV = 98%). The weighting code in brackets is a 13-digit code corresponding to the 13 features in the S4TE scheme and is reported in the head of the compilation file and the results file (Supplementary Figures S2 and S3).

All parameters in S4TE are relevant for the prediction of L. pneumophila T4Es

Beyond the final outcome of S4TE, we sought to determine the relative importance of each searching feature in the algorithm prediction of *L. pneumophila* T4Es. A variable distribution of the effectors across features was observed; some of them were highly selective and specific, whereas others were less efficient (Table 2). For *Legionella*, we confirmed the importance of the hydrophilic profile in the overall length of the protein and its

Table 2. Enumeration of *L. pneumophila* effectors predicted by individual features implemented in S4TE

S4TE feature	1	2	3	4	5	6	7	8	9	10	11	12	13
True positives	38	223	30	5	21	5	1	96	33	185	107	152	13
False positives	19	48	12	11	23	3	0	31	40	81	28	69	5
PPV (%)	67	82	71	31	48	63	100	76	45	70	79	69	72

The number of true positives (TP), false positives (FP) and the positive predictive value (PPV, expressed in %) is indicated.

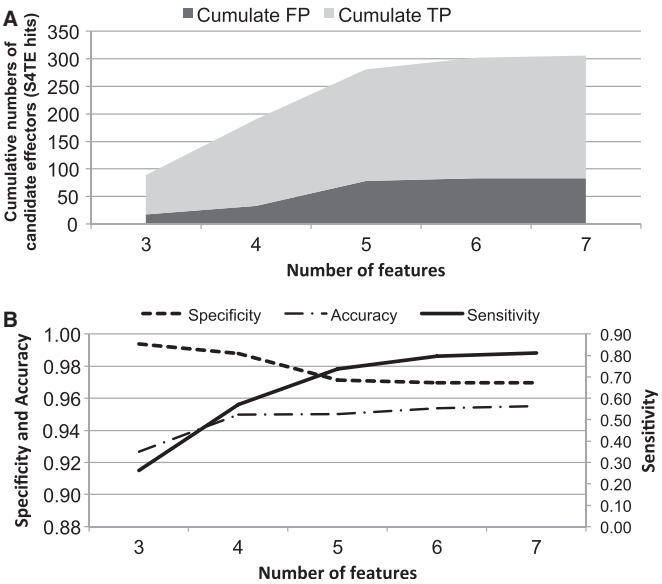


Figure 2. Distribution of the number of features that detected effector candidates in *L. pneumophila*. (**A**) Cumulated numbers of effectors correctly detected (TPs) and called by error (false positives, FP) by S4TE *L. pneumophila* genome. (**B**) Accuracy, sensitivity and specificity of S4TE analysis on *L. pneumophila* genome with combinations of 3, 4, 5, 6 and 7 features.

C-terminus (PPV = 69% for feature 12), the charge of the C-terminus (PPV = 70% for feature 10) and the presence eukaryotic domains (PPV = 71% for feature 3), coiled-coil domains (PPV = 76% for feature 8) and E-block motif (PPV = 72% for feature 13) (Table 2). We then enumerated TP and FP identified by S4TE in *L. pneumophila* according to the number of matching features. TP and FP were well discriminated for combinations of 3 and 4 features (Figure 2A). Even with a slight increase in FP, combinations of 5, 6 and 7 features remained discriminant (Figure 2A). Although accuracy increased from 93% with a combination of 3 features to 95% with 7 features, specificity decreased from 99 to 97% (Figure 2B). The constant rise of the sensitivity from 26% with 3 features to 81% with 7 features shows the importance of our multi-criterion approach to identify a majority of candidate T4Es (Figure 2B). The complete list of feature combinations that generated hits for *L. pneumophila* was used to propose two performance indicators, SI_L and PI_L (see Materials and Methods section and Supplementary Table S2). These indicators are included in the result file appended to each predicted effector and will advise the user on the prediction efficacy of the same combination of features on *L. pneumophila*, thus providing additional help to select the right candidates for further biological evaluation.

Genomic localization of T4Es depends on the G+C content, space clustering analysis and local gene density

The distribution of predicted effectors in the genome of *L. pneumophila* was analysed in detail. We first compared the G+C content of effector-containing regions with the mean G+C content (Figure 3). Effectors were found to be mainly in regions with low G+C content (in green; 60% of the predicted effectors), which agrees with the literature (26). Figure 3 also presents the spatial distribution of predicted effectors and indicates that some genomic regions are enriched in clustered effectors. Such clusters give meaningful information about putative gene co-regulation. The low G+C% content and spatial clustering support the hypothesis that the evolutionary origin of effectors was HGT (53,54). In prokaryotes, the genome architecture refers to the relative position of genetic elements on a genome, including gene order and operons (55). The evolution of genome architecture is largely compelled by specific lifestyles, such as intracellular replication of pathogenic bacteria (55). In some lineages, the active duplication of repeated sequences was associated with adaptation to the host (56), and the presence of insertion elements and transposons is a typical feature of pathogenicity islands (57). We therefore hypothesized that some predicted effectors would be associated with regions of low gene density in the genome of pathogenic bacteria. S4TE serves to visualize the distribution of predicted effectors relative to whole genome architecture (Figure 4). As a case study, we compared the FIRs of predicted effector genes with the architecture of the whole genome of *L. pneumophila* (Figures 4A and B). We found that predicted effectors frequently have both FIRs above the genome median value in *L. pneumophila* genome (Figure 4B and C, Table 3). Although 28% of *L. pneumophila* genes reside in GSRs, this percentage peaks at 52.8% for T4Es genes (162 of the 311 predicted) (Figure 4C). Furthermore, 92.3% of predicted T4Es had at least one FIR longer than the genome median and only 7.7% of the predicted effectors had both FIRs below the genome median (Figure 4B). These observations support the view that plastic genome regions with low gene density frequently harbor pathogenicity genes and may play a role in bacterial adaptation.

S4TE successfully predicts T4Es in other bacterial pathogens

S4TE was used to analyse the genome of four representative α- and γ-proteobacteria. Because our validation

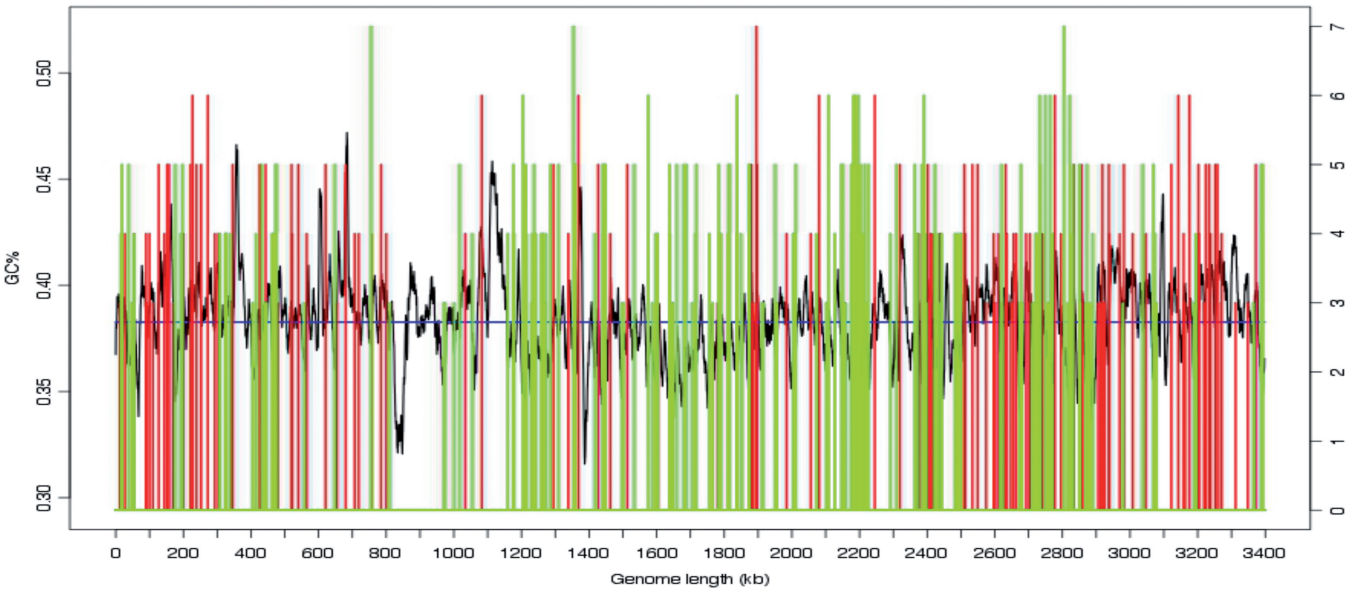


Figure 3. Schematic representation of putative T4 effectors in the *L. pneumophila* genome according to G+C content. This representation is an output file automatically generated by S4TE. The mean GC% is indicated by the blue line. Putative effectors in genomic regions with low G+C content are in green and those in regions with high G+C content are in red.

analysis on *L. pneumophila* suggested that the software had high accuracy (>90%), we next focused on bacterial pathogens of α- and γ-classes having various genome sizes and for which some T4Es were already confirmed experimentally. To evaluate the computational requirements of our algorithm, the different runs on these datasets were timed. It took a maximum of 40 min for the bigger genomes (2–3 Mbp). In the annotation of each genome, S4TE indicated the total number of candidate T4Es, and their position relative to the G+C content and to their FIRs (Table 3). All known T4Es in *Anaplasma marginale* and *B. abortus* were picked up by S4TE, whereas 77% of known T4Es were identified for *C. burnetii* (Table 3). No T4Es of *E. ruminantium* are yet characterized; however, S4TE was able to detect all known T4Es of the closely related *E. chaffeensis* (not shown) and predicted 22 T4Es for *E. ruminantium* (Supplementary Table S3). As expected, orthologs of known T4Es of Anaplasmataceae such as AnkA and ECH_0825 were identified. Moreover, among other putative T4Es, 48% showed a global hydropathy lower than −200, 68% had a C-ter charge >2, 95% had at least three alkaline amino acids in the C-terminal 25 amino acids and 90% harbored characteristic eukaryotic-like domains (Supplementary Table S3). Interestingly, and in contrast to *Legionella*, predicted T4Es in the α-proteobacteria (*E. ruminantium, A. marginale* and *B. abortus*) were mainly localized in genome regions with high G+C content, except for the small chromosome of *B. abortus* (Table 3). However, we noticed a clear exclusion of predicted effectors from GDRs (Table 3). This could be an interesting predicting characteristic to explore in the future. Concerning *C. burnetii*, another γ-proteobacteria-like *Legionella*, the G+C content analysis revealed an equal distribution of predicted effectors throughout the genome but a strong prevalence in GSRs. Although effector prediction by S4TE is still accurate for plasmids (e.g. plasmid of *C. burnetii*) or for small replicons (e.g. chromosome II of *B. abortus*), one has to consider that analyses of the G+C content and genome architecture become meaningless.

DISCUSSION

Before this study, bioinformatics tools for genome-wide annotation of T4Es encoding genes have been developed in a limited number of studies for *L. pneumophila*, *C. burnetii*, *B. abortus* or *A. marginale* (6,10,11,27). However, these *in silico* screening tools search for only few criteria like a combination of homology hits to known effectors and occurrence of eukaryotic-like domains or motifs and coiled-coil domains (10). In another study, candidate T4Es were essentially selected on the basis of their hydropathy profiles and by eliminating proteins with known housekeeping functions and/or with predicted localization signals (11). Finally, Chen *et al.* investigated the genome of *C. burnetii* only for *L. pneumophila* paralogs before experimental validation (6). Our purpose with S4TE was to develop a more complete bioinformatics solution that not only looked for a wider range of sequence features for T4Es prediction but also proposed several visualization interfaces for the outputs. By investigating independently a set of 13 features, combining the different outcomes in a single score and comparing the score with a pre-defined threshold, we provide a computational method that can help biologists in selecting the best potential targets for subsequent experimental validation. S4TE also offers useful services for decision-makers like the representation of genome G+C content and gene density linked to the localization of T4Es and finally performance indicators

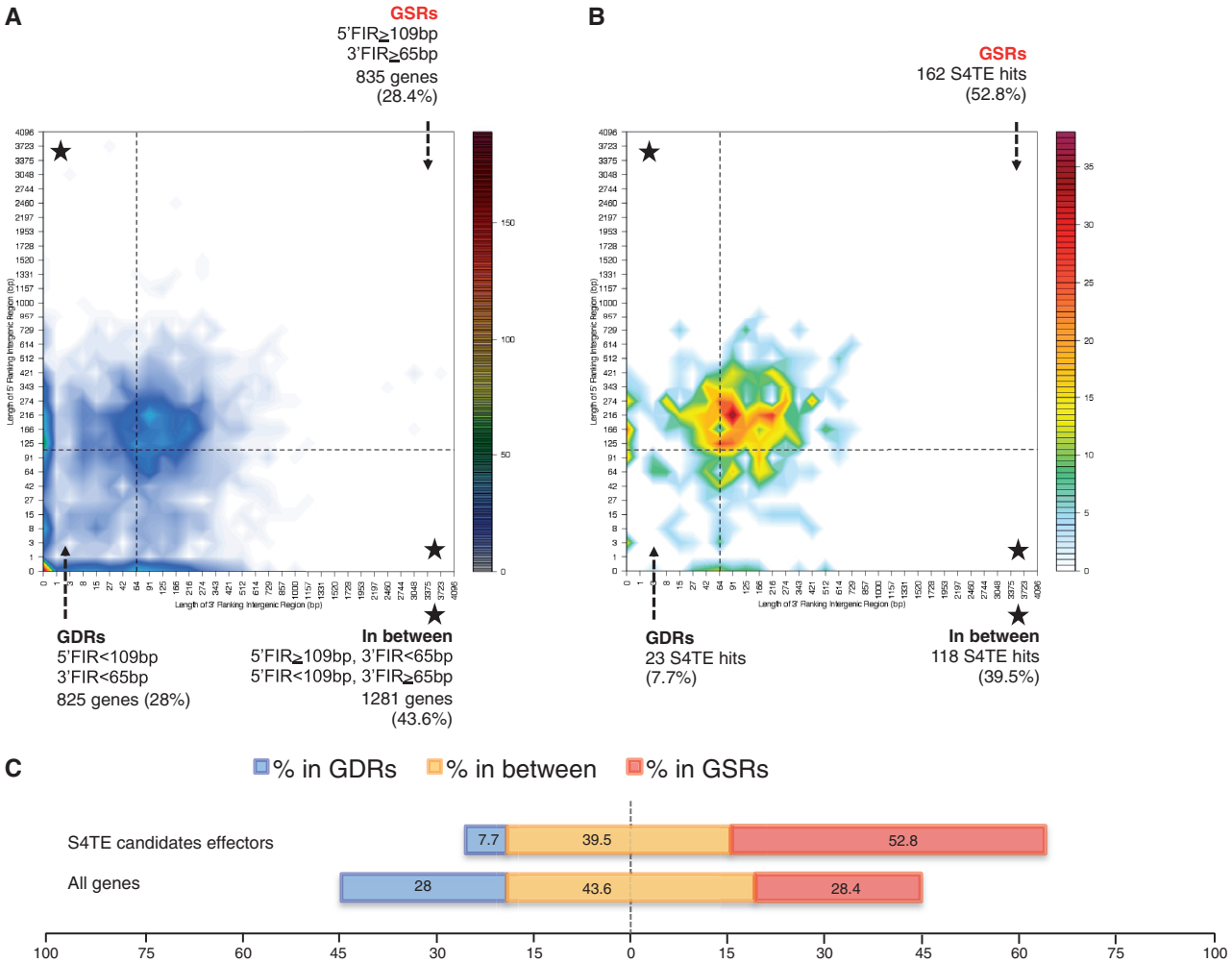


Figure 4. Distribution of *L. pneumophila* genes and predicted T4 effectors according to local gene density (measured as length of flanking intergenic regions, FIRs). (**A**) Distribution of *L. pneumophila* genes according to their FIRs. Genes were sorted in two-dimensional bins according to the length of their 5′ (*y*-axis) and 3′ (*x*-axis) FIR lengths. The number of genes in bins is represented by a color-coded density graph. Genes with both FIRs longer than the median length of FIRs were considered as gene-sparse region (GSR) genes. Genes with both FIRs below the median value were considered as gene-dense region (GDR) genes. In between genes are genes with a long 5′ FIR and short 3′ FIR, and inversely. For *L. pneumophila*, this median value is 109-bp for 5′ FIRs and 65 bp for 3′ FIRs. The dotted line for the median length of FIR delimits the genes in GSR, GDR and in between. (**B**) Distribution of predicted T4 effectors according to their FIRs. The number of hits per T4 effectors in bins is indicated by a color scale. (**C**) Distribution of predicted T4 effectors in the GSRs and GDRs of *L. pneumophila*. The proportion of T4 effectors in GSRs, in between and GDRs is shown in red, yellow and blue, respectively, with percentage indicated.

Table 3. Predicted T4 effectors in various genomes of α- and γ-proteobacteria

Genome	ORF[a]	Known T4Es (%)[b]	Predicted T4Es (%)[c]	Predicted TP (%)[d]	Mean GC (%)[e]	High GC (%)[f]	Low GC (%)[g]	GDRs (%)[h]	IB (%)[i]	GSRs (%)[j]
Ehrlichia ruminantium, strain Gardel	950	NA	22 (2,32)	NA	28	68	32	23	36	41
Anaplasma marginale	963	4 (0,42)	26 (2,70)	100	50	62	38	15	46	38
Brucella abortus chr I	2000	3 (0,15)	53 (2,65)	100	57	62	38	25	40	34
Brucella abortus chr II	1034	1 (0,10)	17 (1,64)	100	57	41	59	6	65	29
Coxiella burnetii	2085	43 (2,06)	126 (6,04)	77	43	50	50	15	41	44
Coxiella burnetii pl	36	1 (2,78)	4 (11,11)	100	39	50	50	25	25	50
Legionella pneumophila	2943	275 (9,34)	311 (10,57)	81	38	40	60	8	40	53

[a]Number of ORFs in the genome.
[b]Number and proportion of known T4 effectors in the genome.
[c]Number and proportion of predicted T4 effectors.
[d]Proportion of true positives in S4TE prediction.
[e]Mean G+C content of the genome.
[f]Proportion of predicted T4 effectors in genomic regions with high G+C content.
[g]Proportion of predicted T4 effectors in genomic regions with low G+C content.
[h]Proportion of predicted T4 effectors in gene-dense regions.
[i]Proportion of predicted T4 effectors in 'in between' regions.
[j]Proportion of predicted T4 effectors in gene-sparse regions.
NA, not applicable; TP, true positives; GDRs, gene dense regions; GSRs, gene sparse regions; T4Es, type IV effectors.

calculated on *L. pneumophila*. The performances of S4TE, which uses 13 different criterions for T4Es prediction and three complementary analyses of genome contents linked to the localization of predicted effectors, were compared with the performances of other algorithms. *C. burnetii* and *L. pneumophila* are exceptional cases for which a broad repertoire of T4 effectors is identified and widely characterized. *In silico* screening with an accurate machine learning prediction algorithm is possible only when sets of negative and positive effector proteins are known, such as for *L. pneumophila* (27). However, for the vast majority of pathogenic bacteria with a T4SS, only few T4 effectors are known. The identification of novel effectors and the subsequent characterization of their function and their targets is a major step to understand how T4SS contributes to bacterial virulence. However, direct biological screening for T4Es can be a tough task, especially for obligate intracellular bacteria that are difficult to cultivate.

In this context, S4TE was designed as an easy-to-use, versatile and customizable algorithm for the prediction of putative effector proteins secreted by the T4SS of proteobacteria whatever the genome size. The high PPV obtained in the large T4Es repertoire of *Legionella* illustrates the relevance of the features combined by S4TE. These features were selected from independent searches on all known T4 effectors on *L. pneumophila*, *C. burnetii*, *A. tumefaciens*, *B. abortus*, *Bartonella* spp., *Anaplasma* spp. and *Ehrlichia* spp. The strength of S4TE relies on the compilation of these features to find TPs. However, user awareness must be raised of the fact that S4TE remains only a useful step toward the identification of T4Es, as non-effector proteins with characteristics similar to T4Es can also be selected, resulting in false positives. Experimental validation of T4SS-dependent translocation is therefore required to establish the effector status of the predicted proteins. Given that the S4TE algorithm is based on characteristics of known T4Es from different bacterial species, genera and even classes, we showed that it might be applicable to other distant pathogenic bacteria. This agrees with a growing number of studies suggesting conserved mode of action or targets for effectors across bacteria classes (21,24,58). The high number of true effectors picked with a potential C-terminal secretion signal seems in line with a T4SS biological function. Also, a previous study has localized the secretion signal at the effector C-terminus of T4Es (38).

Regarding *E. ruminantium,* our main study model, S4TE was able to identify 22 putative T4SS substrates that may contribute to modulation or evasion of host cellular processes. However, further biological testing of their T4SS-dependent secretion is required. Predicting the function of the identified *Ehrlichia* T4SS substrates based solely on their domains is a difficult task. Functional characterization of these candidate T4Es will provide valuable information about the molecular mechanisms underlying the pathogenesis of *Ehrlichia*. Finally, an integrated comprehension of the regulation of T4SS expression and translocation events during infection of host cells will establish the interaction of *Ehrlichia* with its environment.

Future directions

Depending on the availability of biologically validated effectors in α-proteobacteria and confirmation of a strong link between T4Es positions in the genome and the density of genes at these positions, the dense/sparse-gene feature will be integrated in the S4TE algorithm as a predictive value, as done previously for plant pathogenic fungi and oomycetes (51,52,59). We could also refine cutoffs and weighting, as well as add new features when new effectors are discovered. Finally, our approach could be applicable in the identification of candidate effectors in other pathosystems dealing with eukaryotic cells.

CONCLUSION

We have developed a computational tool, S4TE, dedicated to the prediction of candidate bacterial T4 effectors. Our software was designed to identify T4SS effector proteins in α- and γ-proteobacteria. First, the evaluation of S4TE performances demonstrated that the algorithm has a high specificity and high PPVs and NPVs for T4Es. Second, S4TE is time-efficient. Third, S4TE has a very high NPV by default. Yet, a few adjustments can be made by the user to improve confidence in the outcomes. The future-validated *bona fide* T4 effectors will help to refine the S4TE algorithm. In addition, we provide an automated pipeline to analyse effector space clustering and distribution in the genome according to the G+C content and local gene density. The algorithm can be used with default settings, but manual adjustment of the parameters is available. S4TE will be updated when new information (e.g. new validated effectors, new functional domains and new bacterial genomes) becomes available.

S4TE has been registered with the Agency for the Protection of Programs for version 1.0 under registration number IDDN.FR.001.310023.000.S.P.2012.000.31230, filed in June 2012.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [60].

ACKNOWLEDGEMENTS

The authors thank T. Lefrançois and D. Pleydell for useful suggestions, L. Bournez for her aid with Figure 1 and S. Gonzalez-Rizzo for her support and fruitful discussions. D.F.M also thanks P.L. and M.J. Meyer for inspiration. The authors are grateful to G. Burkhart and anonymous reviewers for critical reading and improvement of the manuscript.

FUNDING

UE, FEDER 2007–2013 [FED 1/1.4-30305] 'Risk in animal and plant health'. Funding for open access charge: UE, FEDER 2007–2013 [FED 1/1.4-30305] 'Risk in animal and plant health'.

Conflict of interest statement. None declared.

REFERENCES

1. Hicks,S.W. and Galan,J.E. (2013) Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. *Nat. Rev. Micro.*, **11**, 316–326.
2. Vergunst,A.C., Schrammeijer,B., den Dulk-Ras,A., de Vlaam,C.M., Regensburg-Tuïnk,T.J. and Hooykaas,P.J. (2000) VirB/D4-dependent protein translocation from Agrobacterium into plant cells. *Science*, **290**, 979–982.
3. Vergunst,A.C., van Lier,M.C., den Dulk-Ras,A., Stüve,T.A., Ouwehand,A. and Hooykaas,P.J. (2005) Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of Agrobacterium. *Proc. Natl Acad. Sci. USA*, **102**, 832–837.
4. Schulein,R., Guye,P., Rhomberg,T.A., Schmid,M.C., Schröder,G., Vergunst,A.C., Carena,I. and Dehio,C. (2005) A bipartite signal mediates the transfer of type IV secretion substrates of Bartonella henselae into human cells. *Proc. Natl Acad. Sci. USA*, **102**, 856–861.
5. Backert,S. and Meyer,T.F. (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr. Opin. Microbiol.*, **9**, 207–217.
6. Chen,C., Banga,S., Mertens,K., Weber,M.M., Gorbaslieva,I., Tan,Y., Luo,Z.Q. and Samuel,J.E. (2010) Large-scale identification and translocation of type IV secretion substrates by *Coxiella burnetii*. *Proc. Natl Acad. Sci. USA*, **107**, 21755–21760.
7. Rikihisa,Y. and Lin,M. (2010) Anaplasma phagocytophilum and Ehrlichia chaffeensis type IV secretion and Ank proteins. *Curr. Opin. Microbiol.*, **13**, 59–66.
8. Gomez-Valero,L., Rusniok,C., Cazalet,C. and Buchrieser,C. (2011) Comparative and functional genomics of legionella identified eukaryotic like proteins as key players in host-pathogen interactions. *Front. Microbiol.*, **2**, 208.
9. Zhu,W., Banga,S., Tan,Y., Zheng,C., Stephenson,R., Gately,J. and Luo,Z.-Q. (2011) Comprehensive identification of protein substrates of the Dot/Icm type IV transporter of *Legionella pneumophila*. *PLoS One*, **6**, e17638.
10. Marchesini,M.I., Herrmann,C.K., Salcedo,S.P., Gorvel,J.P. and Comerci,D.J. (2011) In search of Brucella abortus type IV secretion substrates: screening and identification of four proteins translocated into host cells through VirB system. *Cell. Microbiol.*, **13**, 1261–1274.
11. Lockwood,S., Voth,D.E., Brayton,K.A., Beare,P.A., Brown,W.C., Heinzen,R.A. and Broschat,S.L. (2011) Identification of Anaplasma marginale type IV secretion system effector proteins. *PLoS One*, **6**, e27724.
12. Liu,H., Bao,W., Lin,M., Niu,H. and Rikihisa,Y. (2012) Ehrlichia type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating host MnSOD. *Cell Microbiol.*, **14**, 1037–1050.
13. Allsopp,B.A. (2010) Natural history of *Ehrlichia ruminantium*. *Vet. Parasitol.*, **167**, 123–135.
14. Dumler,J.S., Barbet,A.F., Bekker,C.P., Dasch,G.A., Palmer,G.H., Ray,S.C., Rikihisa,Y. and Rurangirwa,F.R. (2001) Reorganization of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: unification of some species of *Ehrlichia* with Anaplasma, Cowdria with *Ehrlichia* and *Ehrlichia* with Neorickettsia, descriptions of six new species combinations and designation of *Ehrlichia* equi and 'HGE agent' as subjective synonyms of *Ehrlichia* phagocytophila. *Int. J. Syst. Evol. Microbiol.*, **51**, 2145–2165.
15. Xiong,Q. and Rikihisa,Y. (2011) Subversion of NPC1 pathway of cholesterol transport by Anaplasma phagocytophilum. *Cell Microbiol.*, **14**, 560–576.
16. Niu,H., Xiong,Q., Yamamoto,A., Hayashi-Nishino,M. and Rikihisa,Y. (2012) Autophagosomes induced by a bacterial Beclin 1 binding protein facilitate obligatory intracellular infection. *Proc. Natl Acad. Sci. USA*, **109**, 20800–20807.
17. Caturegli,P., Asanovich,K.M., Walls,J.J., Bakken,J.S., Madigan,J.E., Popov,V.L. and Dumler,J.S. (2000) ankA: an *Ehrlichia* phagocytophila group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect. Immun.*, **68**, 5277–5283.
18. Park,J., Kim,K.J., Choi,K.-S., Grab,D.J. and Dumler,J.S. (2004) Anaplasma phagocytophilum AnkA binds to granulocyte DNA and nuclear proteins. *Cell Microbiol.*, **6**, 743–751.
19. Lin,M., Den Dulk-Ras,A., Hooykaas,P.J. and Rikihisa,Y. (2007) Anaplasma phagocytophilum AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cell. Microbiol.*, **9**, 2644–2657.
20. Garcia-Garcia,J.C., Rennoll-Bankert,K.E., Pelly,S., Milstone,A.M. and Dumler,J.S. (2009) Silencing of host cell CYBB gene expression by the nuclear effector AnkA of the intracellular pathogen Anaplasma phagocytophilum. *Infect. Immun.*, **77**, 2385–2391.
21. Bierne,H. and Cossart,P. (2012) When bacteria target the nucleus: the emerging family of nucleomodulins. *Cell Microbiol.*, **14**, 622–633.
22. Niu,H., Kozjak-Pavlovic,V., Rudel,T. and Rikihisa,Y. (2010) Anaplasma phagocytophilum Ats-1 is imported into host cell mitochondria and interferes with apoptosis induction. *PLoS Pathog.*, **6**, e1000774.
23. Ham,H., Sreelatha,A. and Orth,K. (2011) Manipulation of host membranes by bacterial effectors. *Nat. Rev. Micro.*, **9**, 635–646.
24. Anderson,D.M. and Frank,D.W. (2012) Five mechanisms of manipulation by bacterial effectors: a ubiquitous theme. *PLoS Pathog.*, **8**, e1002823.
25. Ninio,S. and Roy,C.R. (2007) Effector proteins translocated by Legionella pneumophila: strength in numbers. *Trends Microbiol.*, **15**, 372–380.
26. De Felipe,K.S., Glover,R.T., Charpentier,X., Anderson,O.R., Reyes,M., Pericone,C.D. and Shuman,H.A. (2008) Legionella eukaryotic-like type IV substrates interfere with organelle trafficking. *PLoS Pathog.*, **4**, e1000117.
27. Burstein,D., Zusman,T., Degtyar,E., Viner,R., Segal,G. and Pupko,T. (2009) Genome-scale identification of *Legionella pneumophila* effectors using a machine learning approach. *PLoS Pathog.*, **5**, e1000508.
28. Carey,K.L., Newton,H.J., Lührmann,A. and Roy,C.R. (2011) The Coxiella burnetii Dot/Icm system delivers a unique repertoire of type IV effectors into host cells and is required for intracellular replication. *PLoS Pathog.*, **7**, e1002056.
29. Prakash,A., Yogeeshwari,S., Sircar,S. and Agrawal,S. (2011) Protein domain of unknown function 3233 is a translocation domain of autotransporter secretory mechanism in gamma proteobacteria. *PLoS One*, **6**, e25570.
30. Kalderon,D., Roberts,B.L., Richardson,W.D. and Smith,A.E. (1984) A short amino acid sequence able to specify nuclear location. *Cell*, **39**, 499–509.
31. Nguyen Ba,A.N., Pogoutse,A., Provart,N. and Moses,A.M. (2009) NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, **10**, 202.
32. Claros,M.G. (1995) MitoProt, a Macintosh application for studying mitochondrial proteins. *Comput. Appl. Biosci.*, **11**, 441–447.
33. Al-Quadan,T., Price,C.T., London,N., Schueler-Furman,O. and Abukwaik,Y. (2011) Anchoring of bacterial effectors to host membranes through host-mediated lipidation by prenylation: a common paradigm. *Trends Microbiol.*, **19**, 573–579.
34. Ivanov,S.S., Charron,G., Hang,H.C. and Roy,C.R. (2010) Lipidation by the host prenyltransferase machinery facilitates membrane localization of *Legionella pneumophila* effector proteins. *J. Biol. Chem.*, **285**, 34686–34698.
35. Lupas,A., Van Dyke,M. and Stock,J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
36. Boysen,R.I., Jong,A.J., Wilce,J.A., King,G.F. and Hearn,M.T. (2002) Role of interfacial hydrophobic residues in the stabilization of the leucine zipper structures of the transcription factors c-Fos and c-Jun. *J. Biol. Chem.*, **277**, 23–31.
37. Kyte,J. and Doolittle,R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.
38. Nagai,H., Cambronne,E.D., Kagan,J.C., Amor,J.C., Kahn,R.A. and Roy,C.R. (2005) A C-terminal translocation signal required for Dot/Icm-dependent delivery of the Legionella RalF protein to host cells. *Proc. Natl Acad. Sci. USA*, **102**, 826–831.
39. Huang,L., Boyd,D., Amyot,W.M., Hempstead,A.D., Luo,Z.-Q., O'Connor,T.J., Chen,C., Machner,M., Montminy,T. and

Isberg,R.R. (2011) The E Block motif is associated with *Legionella pneumophila* translocated substrates. *Cell. Microbiol.*, **13**, 227–245.
40. Altman,E. and Segal,G. (2008) The response regulator CpxR directly regulates expression of several *Legionella pneumophila* Icm/Dot components as well as new translocated substrates. *J. Bacteriol.*, **190**, 1985–1996.
41. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
42. Gal-Mor,O., Zusman,T. and Segal,G. (2002) Analysis of DNA regulatory elements required for expression of the *Legionella pneumophilaicm* and dot virulence genes. *J. Bacteriol.*, **184**, 3823–3833.
43. Cunnac,S., Occhialini,A., Barberis,P., Boucher,C. and Genin,S. (2004) Inventory and functional analysis of the large Hrp regulon in Ralstonia solanacearum: identification of novel effector proteins translocated to plant host cells through the type III secretion system. *Mol. Microbiol.*, **53**, 115–128.
44. Johnson,M., Zaretskaya,I., Raytselis,Y., Merezhuk,Y., McGinnis,S. and Madden,T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36**, W5–W9.
45. Punta,M., Coggill,P.C., Eberhardt,R.Y., Mistry,J., Tate,J., Boursnell,C., Pang,N., Forslund,K., Ceric,G., Clements,J. *et al.* (2011) The Pfam protein families database. *Nucleic Acids Res.*, **40**, D290–D301.
46. Barta,M.L., Dickenson,N.E., Patil,M., Keightley,A., Wyckoff,G.J., Picking,W.D., Picking,W.L. and Geisbrecht,B.V. (2012) The structures of coiled-coil domains from type III secretion system translocators reveal homology to pore-forming toxins. *J. Mol. Biol.*, **417**, 395–405.
47. Delahay,R.M. and Frankel,G. (2002) Coiled-coil proteins associated with type III secretion systems: a versatile domain revisited. *Mol. Microbiol.*, **45**, 905–916.
48. Lifshitz,Z., Burstein,D., Peeri,M., Zusman,T., Schwartz,K., Shuman,H.A., Pupko,T. and Segal,G. (2013) Computational modeling and experimental validation of the Legionella and Coxiella virulence-related type-IVB secretion signal. *Proc. Natl Acad. Sci. USA*, **110**, E707–E715.
49. Kirzinger,M.W. and Stavrinides,J. (2012) Host specificity determinants as a genetic continuum. *Trends Microbiol.*, **20**, 88–93.
50. Haas,B.J., Kamoun,S., Zody,M.C., Jiang,R.H., Handsaker,R.E., Cano,L.M., Grabherr,M., Kodira,C.D., Raffaele,S., Torto-Alalibo,T. *et al.* (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature*, **461**, 393–398.
51. Raffaele,S., Farrer,R.A., Cano,L.M., Studholme,D.J., Maclean,D., Thines,M., Jiang,R.H., Zody,M.C., Kunjeti,S.G., Donofrio,N.M. *et al.* (2010) Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science*, **330**, 1540–1543.
52. Raffaele,S., Win,J., Cano,L.M. and Kamoun,S. (2010) Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC Genomics*, **11**, 637.
53. Degtyar,E., Zusman,T., Ehrlich,M. and Segal,G. (2009) A Legionella effector acquired from protozoa is involved in sphingolipids metabolism and is targeted to the host cell mitochondria. *Cell. Microbiol.*, **11**, 1219–1235.
54. Moliner,C., Raoult,D. and Fournier,P.E. (2009) Evidence that the intra-amoebal Legionella drancourtii acquired a sterol reductase gene from eukaryotes. *BMC Res. Notes*, **2**, 51.
55. Koonin,E.V. and Wolf,Y.I. (2010) Constraints and plasticity in genome and molecular-phenome evolution. *Nat. Rev. Genet.*, **11**, 487–498.
56. Collins,N.E., Liebenberg,J., de Villiers,E.P., Brayton,K.A., Louw,E., Pretorius,A., Faber,F.E., van Heerden,H., Josemans,A., van Kleef,M. *et al.* (2005) The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proc. Natl Acad. Sci. USA*, **102**, 838–843.
57. Juhas,M., Crook,D.W. and Hood,D.W. (2008) Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell. Microbiol.*, **10**, 2377–2386.
58. Shames,S.R. and Finlay,B.B. (2012) Bacterial effector interplay: a new way to view effector function. *Trends Microbiol.*, **20**, 214–219.
59. Saunders,D.G., Win,J., Cano,L.M., Szabo,L.J., Kamoun,S. and Raffaele,S. (2012) Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One*, **7**, e29847.
60. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 02 July 2014
doi: 10.3389/fcimb.2014.00086

Understanding *Anaplasmataceae* pathogenesis using "Omics" approaches

Ludovic Pruneau[1,2,3†], Amal Moumène[1,2,3], Damien F. Meyer[1,2], Isabel Marcelino[1,2,4,5], Thierry Lefrançois[1,2] and Nathalie Vachiéry[1,2]*

[1] CIRAD, BIOS, UMR CMAEE, Petit-Bourg, France
[2] INRA, BIOS, UMR CMAEE, Montpellier, France
[3] Université des Antilles et de la Guyane, Pointe-à-Pitre, France
[4] IBET, Apartado, Oeiras, Portugal
[5] ITQB-UNL, Estação Agronómica Nacional, Oeiras, Lisboa, Portugal

Edited by:
Muriel Vayssier-Taussat, INRA, France

Reviewed by:
Ulrike G. Munderloh, University of Minnesota, USA
Bernard Mari, Centre National de la Recherche Scientifique, France
Henri-Jean Boulouis, Ecole Nationale Vétérinaire d'Alfort, France

***Correspondence:**
Nathalie Vachiéry, CIRAD, BIOS, UMR CMAEE, Domaine de Duclos, 97170 Petit Bourg, Guadeloupe, France
e-mail: nathalie.vachiery@cirad.fr

† Present address:
Ludovic Pruneau, Vector-Borne Disease Laboratories, Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Baton Rouge, USA

This paper examines how "Omics" approaches improve our understanding of *Anaplasmataceae* pathogenesis, through a global and integrative strategy to identify genes and proteins involved in biochemical pathways key for pathogen-host-vector interactions. The *Anaplasmataceae* family comprises obligate intracellular bacteria mainly transmitted by arthropods. These bacteria are responsible for major human and animal endemic and emerging infectious diseases with important economic and public health impacts. In order to improve disease control strategies, it is essential to better understand their pathogenesis. Our work focused on four *Anaplasmataceae*, which cause important animal, human and zoonotic diseases: *Anaplasma marginale, A. phagocytophilum, Ehrlichia chaffeensis, and E. ruminantium. Wolbachia* spp. an endosymbiont of arthropods was also included in this review as a model of a non-pathogenic *Anaplasmataceae*. A gap analysis on "Omics" approaches on *Anaplasmataceae* was performed, which highlighted a lack of studies on the genes and proteins involved in the infection of hosts and vectors. Furthermore, most of the studies have been done on the pathogen itself, mainly on infectious free-living forms and rarely on intracellular forms. In order to perform a transcriptomic analysis of the intracellular stage of development, researchers developed methods to enrich bacterial transcripts from infected cells. These methods are described in this paper. Bacterial genes encoding outer membrane proteins, post-translational modifications, eukaryotic repeated motif proteins, proteins involved in osmotic and oxidative stress and hypothetical proteins have been identified to play a key role in *Anaplasmataceae* pathogenesis. Further investigations on the function of these outer membrane proteins and hypothetical proteins will be essential to confirm their role in the pathogenesis. Our work underlines the need for further studies in this domain and on host and vector responses to infection.

Keywords: ***Anaplasmataceae, Anaplasma, Ehrlichia,* proteomics, transcriptomics, pathogenesis**

INTRODUCTION

Understanding of *Anaplasmataceae* biology and pathogenesis has been greatly hampered by their obligatory intracellular characteristic, resulting in culture constraints and difficulties in studying the genetics of these bacteria. During the last 10 years, development of *in vitro* models and significant technical progress in molecular biology using high throughput methods allowed new knowledge to be garnered about their genome expression.

Herein, we review recent advances in the understanding of the *Anaplasmataceae* pathogenesis, using transcriptomics and proteomics approaches. First, we present the general characteristics of *Anaplasmataceae* and then focus on the most studied *Anaplasmataceae* bacteria using "Omics" approaches.

ANAPLASMATACEAE AND THEIR ASSOCIATED DISEASES

The *Anaplasmataceae* family belongs to the *Rickettsiales* order, which includes small obligate intracellular α-proteobacteria, most closely related to mitochondria (Merhej and Raoult, 2011). These bacteria are responsible for major endemic and emerging human and animal infectious diseases with important economic and public health impacts. The *Anaplasmataceae* family includes six genera, *Ehrlichia*, *Anaplasma*, *Aegyptianella*, *Wolbachia*, *Neorickettsia* and "*Candidatus* Neoehrlichia" (Dunning Hotopp et al., 2006). These bacteria infect invertebrate hosts that are abundant and ubiquitous in the environment (i.e., ticks, insects, trematodes, nematodes, or mollusks). *Neorickettsia* and *Wolbachia* spp. can be transmitted through generations of invertebrate hosts by both transovarial and trans-stadial transmission,

whereas *Anaplasma* and *Ehrlichia* seem to have only trans-stadial transmission (Stich et al., 1989; Long et al., 2003). All genera except *Wolbachia* are known to infect vertebrates (mammals or birds). The bacteria infect specific host cell types, such as neutrophils, monocytes and macrophages, platelets, erythrocytes or endothelial cells depending on the species. Within the *Anaplasmataceae* family, *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*, which infect humans, have been widely studied.

Table 1 shows host, vector and geographical distributions for *A. marginale, A. phagocytophilum, E. chaffeensis, E. ruminantium*, and *Wolbachia* spp. *A. phagocytophilum* is responsible for anaplasmosis and infects deer, dogs, cats, horses, ruminants, rodents, and humans, inducing human granulocytic anaplasmosis. *E. chaffeensis* infects deer, dogs, and humans, inducing canine ehrlichiosis and human monocytic ehrlichiosis; whereas *A. marginale* and *E. ruminantium* infect only domestic and wild ruminants. *A. marginale* and *E. ruminantium* cause bovine anaplasmosis and heartwater, respectively. Bovine anaplasmosis has the widest distribution among tick-borne diseases. The acute form of the disease induces fever, anemia, weight loss and often death. After infection, animals are asymptomatic carriers and constitute a reservoir for the transmission of the disease. Heartwater is present in Sub-Saharan Africa, the Caribbean and Indian Ocean islands and induces mortality in ruminants and decreases herd productivity. The economic impact of heartwater, associated to mortality and cost treatment (antibiotic and acaricide), is estimated as $46.7 million per year for the Southern community (Vachiéry et al., 2013).

Bacteria from the *Anaplasmataceae* family develop within a cytoplasmic vacuole in the host cell cytoplasm, whereas other members of *Rickettsiales* order escape from the phagosome after entering the host cell and multiply in the cytoplasm before being released in the extracellular environment (**Figure 1**). Different stages of development are defined for *Anaplasmataceae,* which are characterized by their DNA reorganization from dense cored cell (infectious form) to reticulate cell (vegetative form). The reticulated forms multiply by binary fission and form morulae, and then turn into the dense cored cells before being released.

GENERAL FEATURES OF TRANSCRIPTOMIC AND PROTEOMIC STUDIES

Transcriptomics and proteomics describe the complete, or nearly complete, collection of transcripts and proteins of an organism (bacterium, host, or vector) in different conditions. These conditions include the bacterium growth in different arthropod or vertebrate host cells and tissues. Careful analysis of the transcriptome and proteome is essential to understand the functional output of the genome of bacteria, hosts, and vectors (Filiatrault, 2011).

More specifically, in addition to comparative genomics, transcriptomics and proteomics constitute powerful approaches to increase our knowledge on *Anaplasmataceae* pathogenesis. Researchers have the opportunity to study both sides: (i) the functional genomics of *Anaplasmataceae* in host or vector cells and (ii) the functional genomics of host or vector in response to *Anaplasmataceae* infection.

An overview of the different transcriptomic and proteomic studies currently available on the 5 different *Anaplasmataceae* is represented in **Figure 2**. A complete analysis of host and vector transcriptomes in response to bacterial infection as well as the pathogen transcriptome has only been done for *A. phagocytophilum.* Additionally, the pathogen and host proteomes are also available. For *E. ruminantium*, transcriptomic and proteomic studies were performed only on the pathogen side and there are no data on host or vector responses to pathogen infection. Additionally, concerning the bacteria, proteomic and transcriptomic studies were done also on *E. chaffeensis* and *A. marginale*. The use of such "Omics" approaches on bacteria is particularly useful to perform integrative analyses, linking gene and protein expression for these 4 *Anaplasmataceae* (*E. ruminantium, E. chaffeensis, A. phagocytophilum,* and *A. marginale)*. There is a lack of data concerning the proteome of the vector after infection. The effect of *Wolbachia* on its host genes and protein expressions was studied in order to improve our comprehension of this symbiotic interaction, using *Anopheles gambiae* as model.

Globally, most of the functional genomic studies were performed on the pathogen and only preliminary studies on the host or vector sides are available. Thus, there is a strong need

Table 1 | Hosts, vectors, geographical distribution, and associated diseases for selected *Anaplasmataceae* spp.

Species	Host	Principal vector	Disease	Geographical distribution
A. phagocytophilum	Humans, deer, dogs, cats, horses, ruminants, rodents	Ticks *Ixodes* spp.	Human granulocytic anaplasmosis, tick-born fever of ruminants, anaplasmosis	USA, Europe, and Asia
A. marginale	Cattle, wild ruminants	Ticks *Rhicephalus microplus, Dermacentor*	Bovine anaplasmosis	Worldwide in tropical and subtropical regions
E. chaffeensis	Human, deer, dogs	Ticks *Amblyomma americanum*	Human monocytic ehrlichiosis, canine ehrlichiosis	USA, South America, and Asia
E. ruminantium	Cattle, sheep, goats, wild ruminants	Ticks *African Amblyomma* sp. *including variegatum* and *hebraeum*	Heartwater	Sub Saharan Africa, Comoros, Mayotte, Reunion island, Madagascar, Caribbean
Wolbachia spp.	Nematod *Brugia malayi*, Arthropod *Anopheles gambiae*	NA	NA	Worldwide distribution

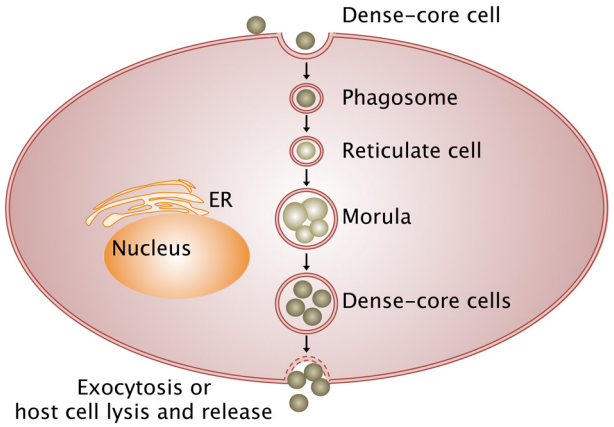


FIGURE 1 | Intracellular life cycle of *Anaplasmataceae*. The *Anaplasmataceae* dense-cored cell develops into replicating reticulate cells in the phagosome, which does not fuse with lysosome and forms morulae (vegetative form). Then, reticulate cells mature into dense-cored cells that are liberated by exocytosis or host cell lysis (infectious form). Arrows, transition between stages; ER, endoplasmic reticulum.

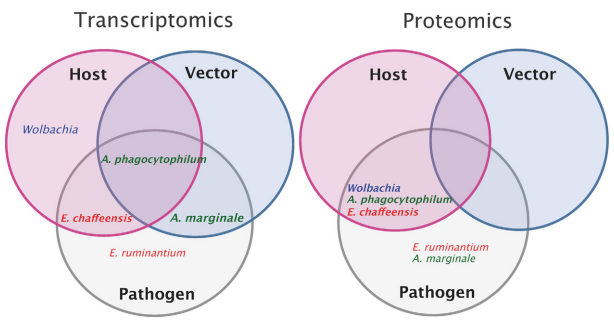


FIGURE 2 | Overview of pathogen, host, and vector transcriptomic and proteomic studies. Transcriptomic or proteomic studies done on host (pink circle) and on vector (blue circle) after pathogen infection (the name of the pathogen is indicated for each study). Transcriptomic or proteomic studies done on the pathogen (gray). Intersections of circles correspond to studies done both on the pathogen and on host or/and vector. In example, transcriptomic studies of *Anaplasma phagocytophilum* and of its host and vector have been done whereas only proteomic studies of *Anaplasma phagocytophilum* and its host have been performed.

to study both vector and host transcriptome and proteome in response to infection in order to better understand pathogen-host and pathogen-vector interactions.

STRATEGIES AND METHODS USED TO PERFORM TRANSCRIPTOMIC AND PROTEOMIC STUDIES ON *ANAPLASMATACEAE*

Various strategies are used to study *Anaplasmataceae* transcriptomic and proteomic studies. Authors generally compare the transcriptome and/or the proteome of (i) bacteria infecting different host and/or vector cells, (ii) bacterial strains with a distinct virulence level and/or from geographically distinct regions, and (iii) bacteria in *in vitro, in vivo,* and *ex vivo* conditions.

Until now, few transcriptomic studies have been performed on the different stages of development of *Anaplasmataceae* members, mainly due to technical constraints associated with the intracellular properties of the bacteria. In one of these studies, the authors compared the *in vitro* gene expression profile between *E. ruminantium* elementary bodies (extracellular dense-cored form) and reticulate bodies (intra-cellular non-infectious forms) (Pruneau et al., 2012). For *Wolbachia* spp., *E. chaffeensis*, *E. ruminantium*, *A. marginale*, and *A. phagocytophilum* transcriptomic and proteomic studies were performed in vertebrate host cells (*in vitro*, *in vivo*, or *ex vivo*). Additional studies were done on the pathogens grown in tick cells for *A. phagocytophilum* and *E. chaffeensis* (transcriptome) and for *E. chaffeensis*, *A. phagocytophilum*, and *A. marginale* (proteome). These different studies allowed further investigation of the interaction between bacteria and its host or vector and the identification of genes, essential for adaptation in host and vector cells.

One of the major constraints in studying the transcriptome and proteome of obligate intracellular bacteria is the large excess of mRNA and proteins of host or vector origin. The removal of eukaryotic RNA and prokaryotic rRNA contaminants is a prerequisite step for *Anaplasmataceae* transcriptomic studies. The depletion of eukaryotic RNA contaminant was achieved mostly through the use of substractive hybridization technologies. The classical approach developed by Ambion (MicrobEnrich Kit), consists of capturing oligonucleotides coupled with magnetic beads that hybridize to the 18S and 28S ribosomal RNA and the poly-adenylated 3′ tail of eukaryotic mRNA. An additional method, known as MicrobExpress (Ambion), uses similar magnetic beads and allows the enrichment of bacterial mRNA by capturing ribosomal and transfer RNA (La et al., 2007). More recently, Emboule and co-workers developed an alternative method for *E. ruminantium*. This method allows both the removal of the eukaryotic RNA and prokaryotic rRNA contaminants and mRNA enrichment from *E. ruminantium* origin (Emboule et al., 2009) and could be adapted for other *Anaplasmataceae*.

For proteomics, various methods have been used to enrich the amount of intracellular bacteria proteins. For example, biochemical fractionation based on differential density and size, or based on harsh detergent treatments dissolving differentially host cells, fractionation based on flow cytometry sorting, or a combination of any of these methods has been used. Purity and yield of the enriched bacteria samples can vary; but generally, dozens to hundreds of bacteria proteins could be successfully detected (Marcelino et al., 2012).

TRANSCRIPTOMES AND PROTEOMES OF *ANAPLASMATACEAE*

Functional genomics studies of *Anaplasmataceae* are essential to identify the main groups of genes regulated during infection of the host or vector. According to their Clusters of Orthologous genes (COG), these genes are classified into the following categories: cell wall membrane biogenesis (including outer membrane proteins); translation, replication and post-translational modifications (PTM); amino acids biosynthesis, metabolism and transport; energy production; intracellular trafficking and secretion

and genes encoding enzymes counteracting osmotic and oxidative stress; and the largest category comprising genes encoding hypothetical proteins.

EXPRESSION OF OUTER MEMBRANE PROTEINS (OMP) DURING *ANAPLASMATACEAE* INFECTION

OMPs play several important roles in bacteria, allowing them to adapt to different environments and host niches. These roles include biogenesis and integrity of the outer membrane, non-specific porin activity, adherence, and membrane associated enzymatic activity (Lin et al., 2002). Some OMPs are porins that form channels, allowing the transport of molecules across lipid bilayer membranes and play a major role in host-interaction. They contribute to nutrient transport, antimicrobial resistance and response to osmotic stress, and are essential for bacteria (Achouak et al., 2001).

OMPs bacteria seem to be crucial in obligate intracellular and could facilitate early interactions with the host and vector cells. They are detected in the proteome of the 5 *Anaplasmataceae* studied, and are the dominant proteins detected in the global proteome of these bacteria. For example, among the 113 *p44* paralogous genes in *A. phagocytophilum*, 110 of them were expressed in the human promyelocytic leukemia cell line, HL-60 (Lin et al., 2011). The first proteome of *E. ruminantium* elementary bodies showed that MAP1 protein was found to be the most predominant protein and seems to be organized as a porin (Marcelino et al., 2012). Other OMP such as MAP 1-6, MAP 1-14, and MAP 1+1 were also identified. Pruneau and co-workers showed an up-regulation of *map1-6* gene expression in *E. ruminantium* reticulate bodies (intracellular forms) compared to elementary bodies (free-living form) (Pruneau et al., 2012). The MAP protein family seems to be essential for *E. ruminantium* intracellular survival; however, their functions remain unknown until now and should be investigated.

DIFFERENTIAL EXPRESSION OF OMP IN INFECTED TICK AND HOST CELLS

In *A. phagocytophilum*, the locus p44, which encodes for several OMPs was expressed in human cell lines and not in tick cell lines. Furthermore, the authors showed a higher proportion of up-regulated genes encoding other OMPs in *A. phagocytophilum* infecting human cell lines rather than in tick cell lines (Nelson et al., 2008). For *E. ruminantium*, MAP 1-1 was expressed only in tick cells and not in host cells (Postigo et al., 2008). A differential expression of OMP-family genes between host cells and vector cells was also observed for *E. chaffeensis* (Kuriakose et al., 2011). These data showed that the OMPs of *Anaplasmataceae* are strongly regulated depending on host or vector cell types and could be important for the development within host and vector.

It would be interesting to better understand the role of these OMPs, which are at the interface between the pathogen, the host, and vector and can represent potential vaccines.

DIFFERENTIAL EXPRESSION OF GENES INVOLVED IN PTM OF BACTERIAL PROTEINS IN INFECTED TICK AND HOST CELLS

Anaplasmataceae with up-regulation of genes leading to PTM can modulate host stress responses and escape from immune system recognition. For *E. chaffeensis*, Kuriakose and co-workers found genes involved in PTM, differentially expressed between host and vector cells, and showed a post-transcriptional regulation of certain *E. chaffeensis* genes involved in host-pathogen-vector interactions (Kuriakose et al., 2011). It would be interesting to further study PTMs on bacterial proteins due to their potential impact on pathogenesis.

DIFFERENTIAL EXPRESSION OF TYPE IV SECRETION SYSTEMS (T4SS) DURING *ANAPLASMATACEAE* INFECTION

All *Anaplasmataceae* have a T4SS, which consists of a multiprotein complex that injects effector proteins into eukaryotic cells. The paradigm of T4SS is that of *Agrobacterium tumefaciens*, which contains 12 *virB/D* genes. Except for *virB1* and *virB5,* all the components of the *A. tumefaciens* T4SS are conserved in *Anaplasmataceae*, but can be duplicated and scattered in several gene clusters (Gillespie et al., 2010). The crucial role of T4SS and its effector proteins in pathogenesis has already been shown for *A. phagocytophilum* and *E. chaffeensis* (Rikihisa and Lin, 2010).

In a recent study, global proteomes of *A. phagocytophilum* and *E. chaffeensis* in human promyelocytic leukemia cell line, HL-60, were characterized (Lin et al., 2011). Proteins of T4SS such as VirB4 and VirD4 were detected as well as the T4SS effectors, AnkA and Ats-1. T4SS components, namely VirB9 (basal body) and VirB11 (ATPase), were detected by proteomic analysis of *E. ruminantium* (Marcelino et al., 2012). For *A. phagocytophilum*, the up-regulation of T4SS genes had already been demonstrated in 3 cell lines: in human cells (HL-60 and HMEC-1) and in tick cells (ISE6) (Nelson et al., 2008). Surprisingly, the *virB2* paralogs were differentially transcribed between the human and tick cells. This result can reflect a specific use of T4SS components depending on the host. Moreover, the effector AnkA was strongly transcribed in HMEC-1, less transcribed in HL-60, and only marginally transcribed in ISE6 (Nelson et al., 2008). In another transcriptomic study, *ankA* was also identified by RNA-sequencing in *Ixodes scapularis* tick salivary glands infected with *A. phagocytophilum* (Mastronunzio et al., 2012). The AnkA effector protein modulates the expression of some host genes and seems to be crucial for *A. phagocytophilum* infection in host cells (Lin et al., 2007). The recent development of bioinformatics software for the identification of candidate T4SS effectors should facilitate the characterization of the role of these proteins of virulence in *Anaplasmataceae* pathogenesis (Meyer et al., 2013). Several components of T4SS were also detected in the global proteome of endosymbiont *Wolbachia* (Bennuru et al., 2011). This suggests a potential role of the T4SS during the endosymbiotic interaction.

EXPRESSION OF BACTERIAL PROTEINS TO CURB OSMOTIC AND OXIDATIVE HOST RESPONSE

For *E. ruminantium*, genes encoding thioredoxin, *trx*, were found to be over-expressed at the reticulate body stage compared to the infectious free stage (Pruneau et al., 2012). Oxidative stress is part of innate immune response and results in a production of reactive oxygen species (ROS) from host cells, which degrade the bacterial membrane. The induction of anti-oxidant enzymes such as thioredoxin or superoxide dismutase diminishes the ROS activity. Other genes involved in the defense against oxidative stress were also up-regulated for *E. chaffeensis* (ECH_0493

encodes for superoxide dismutase) infecting mammalian cells as compared with infected vector cells (Kuriakose et al., 2011) and for *A. phagocytophilum* (APH_0795 encodes for antioxidant AhpC/Tsa family) (Nelson et al., 2008). Superoxide dismutase was also detected in the global proteome of *E. chaffeensis* (Seo et al., 2008). In *E. ruminantium* proteome, TsaA and ElbB proteins, both involved in cell redox homeostasis, were detected. Interestingly, among the several bacterial species compared, the authors found that ElbB was exclusively detected in elementary bodies of *E. ruminantium* (Marcelino et al., 2012). In host and vector cells, *Anaplasmataceae* also fight against osmotic stress by up-regulation of proline-betaine transporter (*proP*). Proline and betaine are two osmoprotectants, which help reduce the hyperosmolarity. These results, showing the fight against the first defense mechanisms of cells, reflect the successful adaptation of *Anaplasmataceae* to their host cells.

IDENTIFICATION OF PATHOGEN METABOLIC ACTIVITIES

Functional genomics of the five *Anaplasmataceae*, focus of this review, revealed the presence of genes/proteins involved in metabolic pathways in host and vector cells. Indeed, many over-expressed genes or proteins were involved either in (i) energy production and conversion and (ii) the transport and metabolism of nucleotides, amino acids, inorganic ions, carbohydrates, and coenzymes.

For *E. ruminantium*, the transcriptome study comparing different stages of development showed that genes involved in the carbohydrate, amino acid, inorganic ion, nucleotide, and coenzyme transports and metabolisms were differentially expressed at both reticulate and elementary body stages. These results suggest that elementary bodies of *E. ruminantium* could be metabolically active (Pruneau et al., 2012). Moreover, at *E. ruminantium* elementary body stage, the majority of proteins detected were related to energy and general metabolism (Marcelino et al., 2012). Lin and co-workers analyzed the global proteome of *A. phagocytophilum* and *E. chaffeensis*. They identified many proteins involved in nucleotide, vitamin, and cofactor biosynthetic pathways (Lin et al., 2011). The comparison between *E. chaffeensis* transcriptome in human cells and in tick cells revealed a larger number of genes with high expression levels in the tick cells. These genes encode for proteins involved in energy production and conversion, nutrient transport, metabolism, cellular process, and translation. The up-regulation of these genes reveals that *E. chaffeensis* has higher metabolic activity in the vector cells than in mammalian cells (Kuriakose et al., 2011).

IDENTIFICATION OF IMPORTANT HYPOTHETICAL PROTEINS IN THE PATHOGENESIS

Regardless of the adopted strategy in transcriptome or proteome studies, genes encoding hypothetical proteins were one of the most represented COG. Kuriakose and co-workers compared the *E. chaffeensis* transcriptome in mammalian *vs.* arthropod hosts (Kuriakose et al., 2011). They showed that genes encoding hypothetical proteins were the most up-regulated genes in human and tick cells. Furthermore, among these genes, they identified 11 highly expressed in human cells that were not expressed in tick cells and 18 expressed in tick cells and not in human cells. These genes do not have any orthologs in other *Ehrlichia* spp. and seem to be required for adaptation and survival in host and vector cells (Kuriakose et al., 2011). Further studies are needed to characterize the function of these genes encoding hypothetical proteins and their role in adaptation and survival in host or vector. Recently, two studies allowed the characterization of the function of the hypothetical protein APH_1235 in *A. phagocytophilum*. The gene encoding APH_1235 was found to be up-regulated in dense cored form (Troese et al., 2011) and the blocking of protein APH_1235 with antibodies reduced infection levels in mammalian cells (Mastronunzio et al., 2012). This result indicates that APH_1235 is required for host cell infection. This protein has homologs in other *Anaplasma* spp. and *Ehrlichia* spp. but not in other bacteria. In *E. ruminantium*, CDS_00640, ortholog of APH_1235, was also up-regulated in the infectious elementary bodies similar to the dense cored form in *A. phagocytophilum* (Pruneau et al., 2012). Even if gene homology analysis is the first step to characterize unknown genes, further experiments will be useful to validate the role of this gene in host cell infection by *E. ruminantium*.

TRANSCRIPTOME AND PROTEOME OF INFECTED HOST AND TICKS

FUNCTIONAL GENOMICS OF HOST CELLS IN RESPONSE TO INFECTION

Regarding functional genomics of host cells in response to infection with *Anaplasmataceae* members, up-regulated genes are generally involved in defense mechanisms and immune responses, such as genes encoding interferon, cytokines, chemokines, and their receptors. Many of these genes were over-expressed in host cells infected with *A. phagocytophilum* (Lee et al., 2008) and *E. chaffeensis* (Miura and Rikihisa, 2009). These genes encode for proteins essential in the first line of defense against bacteria. For *A. phagocytophilum*, the up-regulation of these genes was observed at the early stages post-infection, most of them being down-regulated at later stages post-infection. It seems that *A. phagocytophilum* modulates the expression of these genes during infection to survive in host cells (Lee et al., 2008). In another study, gene expression analyses, using microarrays and real time RT-PCR, were performed in naturally and experimentally infected pigs with *A. phagocytophilum*. These analyses revealed the up-regulation of immune response genes, such as interleukin 1 receptor accessory protein-like 1 (*IL1RAPL1*), T-cell receptor alpha chain (*TCR-alpha*), thrombospondin 4 (*TSP-4*), and Gap junction protein alpha 1 (*GJA1*) genes. These results suggest that pigs control the bacterial infection, particularly through activation of innate immune responses, phagocytosis, and autophagy (Galindo et al., 2012).

Apoptotic pathways are inhibited due to the up-regulation of anti-apoptotic genes and down-regulation of pro-apoptotic genes. This was observed for *A. phagocytophilum* (Lee et al., 2008) and *E. chaffeensis* (Miura and Rikihisa, 2009).

In the study by Miura and Rikihisa, liver cell transcriptome was studied in response to infection with the three strains of *E. chaffeensis*: Wakulla, Liberty, and Arkansas, which induce different histopathologic lesions in liver tissue. The expression profiles of cytokines, chemokines, and their receptors were found to be different among the three strains and could be therefore related to the distinct histopathological lesions (Miura and Rikihisa, 2009).

FUNCTIONAL GENOMICS OF VECTOR CELLS IN RESPONSE TO INFECTION

Concerning vector response to *Anaplasmataceae* infection, only 3 transcriptomics studies have been performed. In the first study, the authors compared *Ixodes scapularis* transcriptome in response to infection with *A. phagocytophilum* and *A. marginale* (Zivkovic et al., 2009). Gene expression profiles were found to be different between the 2 *Anaplasma* species. This difference may reflect the difference of *Anaplasma* spp. developmental cycle in tick cells, or may be due to the fact that *I. scapularis* is not a natural vector of *A. marginale* (Zivkovic et al., 2009). In the second and third studies, the authors studied the transcriptome of vector *Rhipicephalus microplus* infected with *A. marginale* (Zivkovic et al., 2010; Mercado-Curiel et al., 2011). They showed that few genes were regulated in *R. microplus* salivary glands after infection with *A. marginale* (Zivkovic et al., 2010). There was a limited impact of *A. marginale* infection on the tick gene expression compared to uninfected ticks, suggesting minor effects on tick activity or survival. The differentially expressed genes encoding putative proteins are involved in binding, catalytic/enzymatic activity, transport, DNA/RNA metabolism, and structural molecules. Among them, three genes putative von Willebrand factor, a flagelliform silk protein, and subolesin genes, seem to be important for infection and multiplication of *A. marginale* in *R. microplus*. In the third study, the authors showed a differential gene expression between the *R. microplus* midgut and salivary gland in response to feeding and regulation in the salivary gland over a period of time (Mercado-Curiel et al., 2011). This result illustrates the need to study the pathogen-vector interaction during the feeding process.

CONCLUSION AND FUTURE PERSPECTIVES

Anaplasmataceae organisms must be very versatile to survive in different microenvironments. During their extracellular and infective stage, they need to escape the immune system. Inside the host/vector cells, they need to counteract the innate immune response and to subvert cellular processes to survive and/or replicate. Functional genomics facilitate the study of the modulation of genes and proteins expression depending on environmental conditions. A first high-throughput screening of gene/protein expression by "Omics" approaches on the pathogen/host/vector allows a general view of differentially expressed genes/proteins. The second step involves the focusing on specific mechanisms, gene function and pathways of interest to go further in the comprehension of the pathogen biology and pathogenesis using classical approaches.

Resistance to host innate defense mechanisms, like osmotic stress and oxidative burst, are identified in several studies and these mechanisms seem to play a key role for the *Anaplasmataceae* intracellular development. In addition, OMPs and PTMs seem to have an important role in the pathogenesis.

The majority of transcripts and proteins identified by these studies are of unknown functions. They probably have an important role in pathogenesis and should be investigated. The generation of knockout mutants for genes of interest is now technically feasible in *Anaplasmataceae*. The recent development of transient and stable *in vitro* transfection systems for *A. marginale* (Felsheim et al., 2010; Noh et al., 2011) and the optimization of random and targeted mutagenesis for *E. chaffeensis* (Cheng et al., 2013) pave the way for further promising functional analyses of these bacteria. For example, in a recent study, the authors compared the global transcriptome of transformed *A. marginale* strain *vs.* wild type and identified candidate genes for the development of slow growing attenuated vaccines (Pierle et al., 2013). The development of reliable high throughput RNA sequencing methods will replace the use of microarrays for the better understanding of bacterial biology. This could lead to the deciphering of the role of non-coding RNAs, which have not been studied in *Anaplasmataceae*. They could be involved in the regulation of key processes and could have a major role in the pathogenesis of these bacteria.

In the past decade, significant efforts in improving analytical technologies related to measuring mRNA, proteins, and metabolites have been made. Nowadays, new breakthroughs in host-pathogen-vector research rely on the development of novel "Omics" approaches that incorporate high throughput sequencing or separation technologies, and other less known "Omics" such as metabolomics, immunomics, and vaccinomics (Bagnoli et al., 2011). In the future, integrated "Omics" investigation of various cellular molecules and their interaction in cells (i.e., interactomes) could lead to a quantified description of cellular metabolism for further hypothesis-driven investigation. Those efforts will probably lead to fundamentally new insights into bacterial metabolism during intracellular development.

AUTHOR CONTRIBUTIONS

Ludovic Pruneau: Gathering information and papers and writing the paper; Nathalie Vachiéry: Proposing the subject and writing the paper; Amal Moumène, Damien F. Meyer: Preparation and interpretation of the figures, gathering information and reviewing the paper. Thierry Lefrançois and Isabel Marcelino: reviewing the paper. All authors agreed on outlines and the final version of the paper.

ACKNOWLEDGMENTS

The authors acknowledge the financial support received from European project, FEDER 2007–2013, FED 1/1.4-30305, "Risque en santé animale et végétale" and the Fundação para a Ciência e Tecnologia (FCT, Lisbon, Portugal; contract number PTDC/CVT/114118/2009). Ludovic Pruneau and Amal Moumène acknowledge financial support for their PhD from the European project, FED 1/1.4-30305." Isabel Marcelino acknowledges financial support from the grant SFRH/BPD/45978/2008 from FCT. The authors thank Alison Savage for critical reading and improvement of the manuscript, and also Dr. Patricia D. Bedford, Dr. Jennifer Pradel, and Claire Fritsch for their helpful corrections.

REFERENCES

Achouak, W., Heulin, T., and Pages, J. M. (2001). Multiple facets of bacterial porins. *FEMS Microbiol. Lett.* 199, 1–7. doi: 10.1111/j.1574-6968.2001.tb10642.x

Bagnoli, F., Baudner, B., Mishra, R. P., Bartolini, E., Fiaschi, L., Mariotti, P., et al. (2011). Designing the next generation of vaccines for global public health. *OMICS* 15, 545–566. doi: 10.1089/omi.2010.0127

Bennuru, S., Meng, Z., Ribeiro, J. M., Semnani, R. T., Ghedin, E., Chan, K., et al. (2011). Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9649–9654. doi: 10.1073/pnas.1011481108

Cheng, C., Nair, A. D., Indukuri, V. V., Gong, S., Felsheim, R. F., Jaworski, D., et al. (2013). Targeted and random mutagenesis of *Ehrlichia chaffeensis* for the identification of genes required for *in vivo* infection. *PLoS Pathog.* 9:e1003171. doi: 10.1371/journal.ppat.1003171

Dunning Hotopp, J. C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S. V., Eisen, J. A., et al. (2006). Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet.* 2:e21. doi: 10.1371/journal.pgen.0020021

Emboule, L., Daigle, F., Meyer, D. F., Mari, B., Pinarello, V., Sheikboudou, C., et al. (2009). Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium*. *BMC Mol. Biol.* 10:111. doi: 10.1186/1471-2199-10-111

Felsheim, R. F., Chavez, A. S., Palmer, G. H., Crosby, L., Barbet, A. F., Kurtti, T. J., et al. (2010). Transformation of *Anaplasma marginale*. *Vet. Parasitol.* 167, 167–174. doi: 10.1016/j.vetpar.2009.09.018

Filiatrault, M. J. (2011). Progress in prokaryotic transcriptomics. *Curr. Opin. Microbiol.* 14, 579–586. doi: 10.1016/j.mib.2011.07.023

Galindo, R. C., Ayllon, N., Smrdel, K. S., Boadella, M., Beltran-Beck, B., Mazariegos, M., et al. (2012). Gene expression profile suggests that pigs (*Sus scrofa*) are susceptible to *Anaplasma phagocytophilum* but control infection. *Parasit. Vectors* 5, 181. doi: 10.1186/1756-3305-5-181

Gillespie, J. J., Brayton, K. A., Williams, K. P., Diaz, M. A., Brown, W. C., Azad, A. F., et al. (2010). Phylogenomics reveals a diverse *Rickettsiales* type IV secretion system. *Infect. Immun.* 78, 1809–1823. doi: 10.1128/IAI.01384-09

Kuriakose, J. A., Miyashiro, S., Luo, T., Zhu, B., and McBride, J. W. (2011). *Ehrlichia chaffeensis* transcriptome in mammalian and arthropod hosts reveals differential gene expression and post transcriptional regulation. *PLoS ONE* 6:e24136. doi: 10.1371/journal.pone.0024136

La, M. V., Francois, P., Rovery, C., Robineau, S., Barbry, P., Schrenzel, J., et al. (2007). Development of a method for recovering *rickettsial* RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. *J. Microbiol. Methods* 71, 292–297. doi: 10.1016/j.mimet.2007.09.017

Lee, H. C., Kioi, M., Han, J., Puri, R. K., and Goodman, J. L. (2008). *Anaplasma phagocytophilum*-induced gene expression in both human neutrophils and HL-60 cells. *Genomics* 92, 144–151. doi: 10.1016/j.ygeno.2008.05.005

Lin, J., Huang, S., and Zhang, Q. (2002). Outer membrane proteins: key players for bacterial adaptation in host niches. *Microbes Infect.* 4, 325–331. doi: 10.1016/S1286-4579(02)01545-9

Lin, M., den Dulk-Ras, A., Hooykaas, P. J., and Rikihisa, Y. (2007). *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cell. Microbiol.* 9, 2644–2657. doi: 10.1111/j.1462-5822.2007.00985.x

Lin, M., Kikuchi, T., Brewer, H. M., Norbeck, A. D., and Rikihisa, Y. (2011). Global proteomic analysis of two tick-borne emerging zoonotic agents: *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. *Front. Microbiol.* 2:24. doi: 10.3389/fmicb.2011.00024

Long, S. W., Zhang, X., Zhang, J., Ruble, R. P., Teel, P., and Yu, X. J. (2003). Evaluation of transovarial transmission and transmissibility of Ehrlichia chaffeensis (*Rickettsiales: Anaplasmataceae*) in *Amblyomma americanum* (*Acari: Ixodidae*). *J. Med. Entomol.* 40, 1000–1004. doi: 10.1603/0022-2585-40.6.1000

Marcelino, I., de Almeida, A. M., Brito, C., Meyer, D. F., Barreto, M., Sheikboudou, C., et al. (2012). Proteomic analyses of *Ehrlichia ruminantium* highlight differential expression of MAP1-family proteins. *Vet. Microbiol.* 156, 305–314. doi: 10.1016/j.vetmic.2011.11.022

Mastronunzio, J. E., Kurscheid, S., and Fikrig, E. (2012). Postgenomic analyses reveal development of infectious *Anaplasma phagocytophilum* during transmission from ticks to mice. *J. Bacteriol.* 194, 2238–2247. doi: 10.1128/JB.06791-11

Mercado-Curiel, R. F., Palmer, G. H., Guerrero, F. D., and Brayton, K. A. (2011). Temporal characterisation of the organ-specific *Rhipicephalus microplus* transcriptional response to *Anaplasma marginale* infection. *Int. J. Parasitol.* 41, 851–860. doi: 10.1016/j.ijpara.2011.03.003

Merhej, V., and Raoult, D. (2011). *Rickettsial* evolution in the light of comparative genomics. *Biol. Rev. Camb. Philos. Soc.* 86, 379–405. doi: 10.1111/j.1469-185X.2010.00151.x

Meyer, D. F., Noroy, C., Mouméne, A., Raffaele, S., Albina, E., and Vachiéry, N. (2013). Searching algorithm for type IV secretion system effectors 1.0: a tool for predicting type IV effectors and exploring their genomic context. *Nucleic Acids Res.* 41, 9218–9229. doi: 10.1093/nar/gkt718

Miura, K., and Rikihisa, Y. (2009). Liver transcriptome profiles associated with strain-specific *Ehrlichia chaffeensis*-induced hepatitis in SCID mice. *Infect. Immun.* 77, 245–254. doi: 10.1128/IAI.00979-08

Nelson, C. M., Herron, M. J., Felsheim, R. F., Schloeder, B. R., Grindle, S. M., Chavez, A. O., et al. (2008). Whole genome transcription profiling of *Anaplasma phagocytophilum* in human and tick host cells by tiling array analysis. *BMC Genomics* 9:364. doi: 10.1186/1471-2164-9-364

Noh, S. M., Ueti, M. W., Palmer, G. H., Munderloh, U. G., Felsheim, R. F., and Brayton, K. A. (2011). Stability and tick transmission phenotype of gfp-transformed *Anaplasma marginale* through a complete *in vivo* infection cycle. *Appl. Environ. Microbiol.* 77, 330–334. doi: 10.1128/AEM.02096-10

Pierle, S. A., Hammac, G. K., Palmer, G. H., and Brayton, K. A. (2013). Transcriptional pathways associated with the slow growth phenotype of transformed *Anaplasma marginale*. *BMC Genomics* 14:272. doi: 10.1186/1471-2164-14-272

Postigo, M., Taoufik, A., Bell-Sakyi, L., Bekker, C. P., de Vries, E., Morrison, W. I., et al. (2008). Host cell-specific protein expression *in vitro* in *Ehrlichia ruminantium*. *Vet. Microbiol.* 128, 136–147. doi: 10.1016/j.vetmic.2007.09.023

Pruneau, L., Emboule, L., Gely, P., Marcelino, I., Mari, B., Pinarello, V., et al. (2012). Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development. *FEMS Immunol. Med. Microbiol.* 64, 66–73. doi: 10.1111/f574-695X.2011.00901.x

Rikihisa, Y., and Lin, M. (2010). *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins. *Curr. Opin. Microbiol.* 13, 59–66. doi: 10.1016/j.mib.2009.12.008

Seo, G. M., Cheng, C., Tomich, J., and Ganta, R. R. (2008). Total, membrane, and immunogenic proteomes of macrophage- and tick cell-derived *Ehrlichia chaffeensis* evaluated by liquid chromatography-tandem mass spectrometry and MALDI-TOF methods. *Infect. Immun.* 76, 4823–4832. doi: 10.1128/IAI.00484-08

Stich, R. W., Kocan, K. M., Palmer, G. H., Ewing, S. A., Hair, J. A., and Barron, S. J. (1989). Transstadial and attempted transovarial transmission of *Anaplasma marginale* by *Dermacentor variabilis*. *Am. J. Vet. Res.* 50, 1377–1380.

Troese, M. J., Kahlon, A., Ragland, S. A., Ottens, A. K., Ojogun, N., Nelson, K. T., et al. (2011). Proteomic analysis of *Anaplasma phagocytophilum* during infection of human myeloid cells identifies a protein that is pronouncedly upregulated on the infectious dense-cored cell. *Infect. Immun.* 79, 4696–4707. doi: 10.1128/IAI.05658-11

Vachiéry, N., Marcelino, I., Martinez, D., and Lefrancois, T. (2013). Opportunities in diagnostic and vaccine approaches to mitigate potential heartwater spreading and impact on the American mainland. *Dev. Biol. (Basel.)* 135, 191–200. doi: 10.1159/000190050

Zivkovic, Z., Blouin, E. F., Manzano-Roman, R., Almazán, C., Naranjo, V., Massung, R. F., et al. (2009). *Anaplasma phagocytophilum* and *Anaplasma marginale* elicit different gene expression responses in cultured tick cells. *Comp. Funct. Genomics* 2009:705034. doi: 10.1155/2009/705034

Zivkovic, Z., Esteves, E., Almazán, C., Daffre, S., Nijhof, A. M., Kocan, K. M., et al. (2010). Differential expression of genes in salivary glands of male Rhipicephalus (*Boophilus*) microplus in response to infection with Anaplasma marginale. *BMC Genomics* 11:186. doi: 10.1186/1471-2164-11-186

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 March 2014; accepted: 10 June 2014; published online: 02 July 2014.
Citation: Pruneau L, Moumène A, Meyer DF, Marcelino I, Lefrançois T and Vachiéry N (2014) Understanding Anaplasmataceae pathogenesis using "Omics" approaches. Front. Cell. Infect. Microbiol. ***4****:86. doi: 10.3389/fcimb.2014.00086*
This article was submitted to the journal Frontiers in Cellular and Infection Microbiology.